Crosstalk, Network Dynamics, and the Evolution of Signaling

By

Michael Allan Rowland

Submitted to the Center for Bioinformatics and the
Graduate Faculty of the University of Kansas
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Committee members

Eric J. Deeds, Ph.D., Chairperson

Wonpil Im, Ph.D.

John Karanicolas, Ph.D.

J. Christian Ray, Ph.D.

Mark Richter, Ph.D.

Date defended: October 03, 2014

The Dissertation Committee for Michael Allan Rowland certifies
that this is the approved version of the following dissertation :

Crosstalk, Network Dynamics, and the Evolution of Signaling

Eric J. Deeds, Ph.D., Chairperson

Date approved: December 09, 2014

Abstract

Cells have developed networks of interacting proteins to process information about their environment and respond appropriately to stimuli. Reversible post-translational modifications alter the functionality of these proteins, transmitting information through the cell. Bacteria primarily utilize Two-Component Signaling (TCS) networks, in which a sensor Histidine Kinase (HK) activates a Response Regulator (RR), which typically acts as a transcription factor. TCS pathways are insulated from one another, each responding to a unique stimulus. In contrast, metazoan signaling networks are extremely complex, to the point that individual pathways are no longer discernible from the web of interactions. Cellular decisions are no longer binary; the overall state of the network determines the response to inputs. In this work, we use mathematical modeling to explore the dynamics that give rise to the dichotomy in network complexity and the evolutionary pressures and benefits of crosstalk, or the lack thereof. We find that proteins can act as competitive inhibitors of each other when competing for a shared enzyme. For example, the phosphorylation of one protein would monopolize a phosphatase, decreasing the concentration of phosphatase available to competing substrates. Consequently, the other substrates would see an increase in their own phosphorylation, indicating the potential for crosstalk mediated by any shared enzyme. The shared competitive inhibition of enzymes by different substrates has a more drastic effect in bacterial TCS pathways. HKs are typically bifunctional, acting as both kinase and phosphatase for their RRs. These dynamics results in a situation in which the introduction of crosstalk to TCS networks would always decrease system efficiency. While the enzymes typical of metazoan networks do not have the same en-

zymatic constraints as TCS networks, the fact that they can evolve crosstalk does not explain the benefits that have driven such complexity. The extensive crosstalk present in metazoans has likely evolved due to the constraints multicellularity has placed on intracellular communication. Because of the complexity of the network, the expression of different signaling components in various cell types results in a high level of diversity in responses to stimuli. Ultimately, our work demonstrates that the cellular context must be considered in interpreting network connectivity.

I dedicate this dissertation to my wife, Kelly, and my daughter, Aurelia.

Without you I would not have been able to come so far.

From this hour I ordain myself loos'd of limits and imaginary lines,

Going where I list, my own master total and absolute,

Listening to others, considering well what they say,

Pausing, searching, receiving, contemplating,

Gently, but with undeniable will, divesting myself of the holds that would hold me.

I inhale great draughts of space,

The east and the west are mine, and the north and the south are mine.

I am larger, better than I thought,

I did not know I held so much goodness.

-Walt Whitman, *Song of the Open Road*

Acknowledgements

There are many people I would like to thank who have supported and guided me up to and through graduate school. I could not have done any of this without you.

First I would like to thank my advisor, Dr. Eric Deeds. He took me into his lab in 2011 with just a semester to prepare for my oral qualifier. Since then he has guided me through many different projects, giving me the freedom to explore my own ideas, knowing when to hold me back, and when to drag me through analyses I did not want to do. Because of him I have gained a new appreciation and vision for signaling network dynamics and an eye for the beauty and power of the written word. He lead me through many of the steps to become a strongly independent, driven scientist.

I am grateful to the members of my committee, Dr. Wonpil Im, Dr. John Karanicolas, Dr. Christian Ray, and Dr. Mark Richter for their time and helpful comments. Additionally, I would like to thank Dr. Ilya Vakser and Dr. Chris Fischer for their comments on my pre-doctoral examination.

The work presented in this dissertation would not be possible without the help from my lab mates and external collaborators. I am grateful for the support and comments from Dr. Walter Fontana as I worked to produce my first project. I would like to thank Joe Greenbaum and Brian Harrison for their contributions to two of my projects. I would also like to thank lab members Ryan Suderman, Zaikun Xu, Dustin Maurer, Addison Schauer, Kerrigan Blake, Adam Smith, and Johnny Israeli for many interest-

ing conversations and ideas.

Dr. Wonpil Im and Dr. Richard Pastor took me on as a student in 2010, before I fully understood the research direction in which I was most interested. It was their support, inspiration, and criticisms that took me from a research technician who was starting to come out of the shadows of his mentors and made the first impressions of what my own scientific research and work ethic should look like. I would like to thank the members of their labs who helped me learn the ropes when I first started at the Center for Bioinformatics, including Dr. Huan Rui, Dr. Taehoon Kim, Dr. Sunhwan Jo, Dr. Kyuil Lee, Dr. Alexander Sodt, Richard Venable, Kevin Song, Danielle Stuhlsatz, and Phillips Morris.

Additionally, I would like to thank my colleagues and friends in the Center for Bioinformatics who have supported me throughout graduate school, including Tammie Powers, Maria Swinger-Inskeep, Jeannette Atkinson, Debbie Douglass-Metsker, Whitney Tran, David Johnson, Dr. Ritesh Kumar, Ragul Gowthaman, Varsha Badal, Ivan Anishchenko, Saveliy Belkin, Xi Cheng, Karen Khar, Jiaqin Li, Andrew Beaven, Dr. Andrea Bazzoli, and Dr. Joanna Slusky.

Dr. David Clemmons, Dr. Ted Busby, and Dr. Laura Maile took a risk and first introduced me to biochemistry research in 2003 when I was a mathematics undergraduate at the University of North Carolina at Chapel Hill. It was through their guidance, trust, and excitement that I gained a love of biology and cell signaling. I co-authored my first articles with them, and it was with their support that I applied and was accepted into graduate school. I would also like to thank my many colleagues from UNC-CH, including Laura Lindsey, Dr. Yashwanth Radhakrishnan, Dr. Xinchun Shen, Dr. Teresa Cascella, Dr. Gang Xi, Dr. Yan Ling, Dr. Tomoko Izawa, Dr. Junyu Ning, Dr. Melissa

Solum, Dr. Katy Gollahon, Dr. Umadevi Veluvolu, Dr. Mijin Kwon, Dr. Jane Badley-Clarke, Dr. Sung-Eun Yoo, Dr. Jaroslava Lieskovska, Dr. Jennifer Sipos, Kevin Sitko, Ariel Aday, Byron Capps, Tiffany Sergent, Lee Allen, Kevin Kohler, Christine Wai, Adam Dreschler, Chris Hanzaker, Paul Dunbar, and the many others with whom I had the pleasure of working alongside.

I would especially like to thank my parents, Dr. Chris and Celia Rowland, who first introduced me to science and mathematics. They have always given me the license to question my world, the information and materials to encourage me to pursue science, and the freedom to make mistakes so long as I learn from them. I would also like to thank my stepmother, Angela Rowland, for her encouragement. I am grateful for my in-laws, Robin and Jim Short, and my stepmother-in-law, Vicky Short, for all of their support and encouragement and continued sense of humor. I would like to thank my siblings Leah and Jason Gardner, Taylor Rowland, Meghan Rowland, and Lyndsay and Michael Ponce. I would also like to express thanks to Laura and Zinn Morton for providing me with room, board, and transportation during my internship in NIH. I would also like to thank Gary and Kellee Pratt for welcoming us to Lawrence and becoming not only family, but friends as well.

I also owe thanks to my many friends in Lawrence. They provided me with a community, inspiration, and support and have seen both me at both my best and worst. In particular, I would like to mention Matt and Emily Cox, Ryan and Tiffany Classen, Brock and Pamela Brown, Palmer and Amanda Davidson, Bob and Kelli Szrot, Anna and Scott Brunk, Jason and Rhonda Dye, Justin and Chelsea Hannon, Chris and Mallory Laing, Tim and Ami Bruce, Casey and Beth Morford, Brian and Angela Heili, Chris and Heather Brust, Chad and Kim Luce, Sue and John Haley, Kevin and Kristianne Jaeger, Brian and Eileen Bangalan, and Craig and Tisha Anderson.

My daughter, Aurelia Rowland, has reminded me that endless patience and love is awarded with the best gifts and to enjoy even the smallest moments of life.

Finally, I would above all like to thank my wife and my best friend, Kelly Rowland. She has always seen the best in me, even when I could not. She saw beyond my messiness and my quirks when we first met. She stabilized me when it became obvious that I could not succeed at life on my own. She encouraged me to apply to graduate school and supported me when I chose to go to the University of Kansas. She worked jobs that may not have been her dream position to support us throughout school. She is the first to comfort me in my lowest lows and is the first to cheer me on in my highest highs. I would not be the man I am without her standing alongside me, equal partners in life. I will love you forever.

Contents

1 Introduction **1**

2 Crosstalk and Competition in Signaling Networks **9**

2.1 Introduction 9

2.2 Materials and Methods 14

2.3 Results 14

2.3.1 1-Kinase/1-Phosphatase loop with two substrates 14

2.3.2 1-Kinase/1-Phosphatase loop with many substrates 18

2.3.3 1-Kinase/2-Phosphatase loop 19

2.3.4 2-Kinase/1-Phosphatase loop 21

2.3.5 Phosphatase tunneling 22

2.3.6 Kinase inhibitors 24

2.4 Discussion 27

3 Phosphatase Specificity and Pathway Insulation in Signaling Networks **30**

3.1 Introduction 30

3.2 Materials and Methods 33

3.3 Results 35

3.3.1 The promiscuity of phosphatases 35

3.3.2 Removing coupling with unsaturatable phosphatases 36

3.3.3 Degradation as a phosphatase substitute 40

3.3.4 Role of phosphatase regulatory subunits in pathway isolation 43
3.4 Discussion 48

4 Crosstalk and the Evolution of Specificity in Two-Component Signaling 51
4.1 Introduction 51
4.2 Materials and Methods 53
4.3 Results 55
4.3.1 Response to changes in RR concentration 55
4.3.2 Competition in TCS 57
4.3.3 Evolutionary trajectories 59
4.3.4 Evidence for near-neutral trajectories 62
4.4 Discussion 66

5 Crosstalk and the Evolvability of Intracellular Communication 69
5.1 Introduction 69
5.2 Materials and Methods 72
5.3 Results 74
5.3.1 Crosstalk and expression provide a diversity of cellular responses 74
5.3.2 Specific expression patterns in human tissues are selected to generate signaling diversity 76
5.3.3 Differential effects of inhibitors across cell types 80
5.4 Discussion 83

6 Conclusion 85

A Appendix for Crosstalk and Competition in Signaling Networks 107
A.1 Systems of Ordinary Differential Equations 107
A.1.1 1-Kinase/1-Phosphatase loop with two substrates 107
A.1.2 1-Kinase/1-Phosphatase loop with many substrates 109

A.1.3 1-Kinase/2-Phosphatase loop . . . 113
A.1.4 2-Kinase/1-Phosphatase loop . . . 115
A.1.5 Cascade with multiple phosphatases . . . 117
A.1.6 Cascade with a single phosphatase . . . 121
A.2 Analytical results for the 1-Kinase/1-Phosphatase loop . . . 124
A.2.1 Mutual inhibition for competitive substrates . . . 124
A.2.2 Steady-state solution for $[S_1^*]$ . . . 125
A.2.3 dS_1^*/dS_2^* is always positive . . . 127
A.3 Analytical results for the 1-Kinase/1-Phosphatase loop with many substrates . . . 131
A.4 Analytical results for the 1-Kinase/2-Phosphatase loop . . . 132
A.4.1 $d[S_2]/d[S_2]_0$ is always positive . . . 133
A.4.2 $dS_1^*/d[S_2]_0$ is always negative . . . 135
A.5 Analytical results for the 2-Kinase/1-Phosphatase loop . . . 136
A.5.1 $d[S_2^*]/d[S_2]_0$ is always positive . . . 137
A.5.2 $dS_1^*/d[S_2]_0$ is always positive . . . 138

B Appendix for Phosphatase Specificity and Pathway Insulation in Signaling Networks 139
B.1 Systems of Ordinary Differential Equations . . . 139
B.1.1 2-Kinase/1-Phosphatase Loop with 2 Substrates . . . 139
B.1.2 2-Kinase/1-Phosphatase Loop with 2 Substrates and 2 Reservoir Proteins . 142
B.1.3 2-Kinase/1-Phosphatase Loop with Many Substrates . . . 146
B.1.4 1-Kinase/1-Substrate Model with Synthesis and Degradation . . . 146
B.1.5 2-Kinase/1-Phosphatase Loop with “Ordered" Phosphatase Adaptors . . . 149
B.1.6 2-Kinase/1-Phosphatase Loop with “Unordered" Phosphatase Adaptors . . 153
B.2 The responses of the substrates of an unsaturated phosphatase are strictly hyperbolic in r . . . 157
B.3 Analytical solution for 1-Kinase/1-Substrate Model with Synthesis and Degradation 159
B.4 UniProt Data . . . 162

C Appendix for Crosstalk and the Evolution of Specificity in Two-Component Signaling 166

C.1 Analytical Results for Competition in TCS . . . 166

C.1.1 Steady-state solution for RR_1^* . . . 166

C.1.2 $r_1\beta_1 - \beta_1 < 0$ . . . 171

C.1.3 Only the positive branch of RR_1^* is realistic . . . 172

C.2 The effects of increased *RR* concentration on TCS . . . 175

C.2.1 $d[RR_1^*]/d[RR_1]_0 > 0$ . . . 175

C.2.2 $dRR_1^*/d[RR_1]_0 < 0$ . . . 176

C.2.3 RR_1^* goes to 0 as $[RR_1]_0 \to \infty$ . . . 178

C.2.4 $\frac{d[RR_1^*]}{d[RR_1]_0}$ goes to 0 as $[RR_1]_0 \to \infty$ . . . 178

C.3 Expanded Model With Inputs and Outputs . . . 181

C.3.1 Details on Parameters for Main Text Figures . . . 188

C.4 Comparison of the Model with Experimental Data . . . 189

C.5 Details of K_A/K_S Analysis . . . 190

C.5.1 Methods . . . 190

C.5.2 Horizontal Gene Transfer . . . 192

C.5.3 Statistical Analyses . . . 194

C.5.4 PAS domain analysis . . . 210

D Appendix for Crosstalk and the Evolvability of Intracellular Communications 213

D.1 Supporting Figures . . . 213

D.2 Methods . . . 216

D.2.1 Evolvable Boolean Networks . . . 216

D.2.2 The Complete KEGG Signaling Network . . . 219

E List of Publications 222

F Permissions 223

List of Figures

1.1 Overview of cell signaling 2
1.2 Signaling pathways 4

2.1 The Goldbeter-Koshland loop 11
2.2 Crosstalk schematic 13
2.3 Results for the 1K1P loop 17
2.4 Influence of phosphatase architecture on network response 20
2.5 Influence of phosphatase tunneling on cascade signals 23
2.6 Effect of kinase inhibitors in the presence of crosstalk 26

3.1 The 2-Kinase/1-Phosphatase Loop. 34
3.2 Removing coupling with unsaturatable phosphatases. 37
3.3 Degradation as a phosphatase. 42
3.4 The "ordered" model of phosphatase regulatory adaptor subunits. 45
3.5 The "unordered" model of phosphatase regulatory adaptor subunits. 47

4.1 TCS pathways vs. Goldbeter-Koshland loops 54
4.2 Effects of competition on TCS signaling 58
4.3 Evolutionary trajectories 61
4.4 Sequence analysis 63

5.1 Bacterial TCS versus human signaling networks 71
5.2 Evolvable Boolean signaling networks 75

5.3 The structure and response diversity of the complete KEGG network 78
5.4 The effects of inhibitors on different cells 81

B.1 The effects of increased phosphatase concentration on substrate crosstalk with strong $K_{M,P,i}$s (*A*) The fraction S_1^* as a function of the concentration of the phosphatase. The concentration of the kinase is increased in order to maintain the values of r_1 and r_2 with constant catalytic rates for different concentrations of the phosphatase. At low concentrations of P, S_1 phosphorylation is increased by activation of the S_2 pathway, moving from $r_2 = 0$ (black) to $r_2 = 2$ (red). As P is expressed in concentrations greater than the substrates, the difference between the curves is removed. However, the fraction S_1^* is greatly increased. Additionally, the fraction of unbound S_1^* decreases with $[P]_0$, indicating that the increase in total fraction S_1^* is likely due to it being bound to the phosphatase. (*B*) The fold increase in S_1^* as a function of the concentration of the phosphatase. As stated above, the crosstalk between S_1 and S_2 is removed when the phosphatase is present in concentrations larger than those of the substrates. 163
B.2 The effects of reservoir proteins on substrate crosstalk (*A*) The fraction S_1^* as a function of r_1. Without the reservoir proteins, S_1 responds to signals from S_2 (red versus black curves). The crosstalk is removed with the addition of the reservoir proteins (orange versus blue curves). Note, however, that the response becomes hyperbolic in r_1. (*B*) The fold increase in S_1^* as a function of reservoir protein concentration. As the concentrations of the reservoir proteins increases, the crosstalk between the substrates is gradually removed. (*C*) The half-life of S_1 phosphorylation as a function of the concentration of reservoir proteins. Note that when the reservoir proteins are at stoichiometric or greater concentrations, the time required to completely dephosphorylate the substrates greatly increases. 163

C.1 Fraction RR^* versus k_f/k_b comparing the analytical solution C.1.3.8 and numerical simulations run to steady state . . . 175
C.2 Concentration vs. Fraction of Phosphorylated RR . . . 180
C.3 Thermodynamic cycle for an input binding an HK . . . 185
C.4 Comparison of the model with experimental data from Batchelor & Goulian, 2003 . 190
C.5 HGT Analysis . . . 195
C.6 K_A/K_S controls for the HK domains . . . 198
C.7 K_A/K_S analysis controls using *Thermotoga maritima* . . . 201
C.8 K_A/K_S analysis controls for the RR interface vs. noninterface . . . 206
C.9 K_A/K_S of the interface vs. K_A/K_S of the noninterface for HK and RR domains . . . 208
C.10 Raw amino acid substitutions . . . 209
C.11 Example alignment from *Halococcus turkmenicus* . . . 209
C.12 K_S of the PAS and K domains of the HKs controls . . . 212

D.1 Node properties versus the expression fraction of a node. (*A*) The indegree of a node versus its expression fraction. The expression fraction of a node is calculated as the ratio of the number of tissues in which the node is expressed to the total number of tissues. There is no correlation between the indegree and the expression fraction (Spearman's ρ = 0.03824749, p = 0.4795). (*B*) The outdegree of a node versus its expression fraction. There is a weak correlation between the outdegree and the expression fraction (Spearman's ρ = 0.15365, p = 4.285 $\times 10^{-3}$), however the data cannot provide a statistically significant linear regression (p = 0.1548, adjusted R-squared = 0.003004). (*C*) The number of shortest paths between an input and an output that includes a node versus its expression fraction. There is a weak correlation between the number of shortest paths and the expression fraction (Spearman's ρ = 0.1506653, p = 5.106 $\times 10^{-3}$), however the data cannot provide a statistically significant linear regression (p = 0.5059, adjusted R-squared = 0.005518). . . . 214

D.2 Kernel density plot of target pleiotropy across the 84 expressed subnetworks. The target pleiotropy is the fraction of the 84 tissues in which inhibiting the node results in a change in output activity. Note that there are no nodes whose inhibition has an effect in all 84 tissues (Target pleiotropy = 1). Additionally, every node can be targeted in order to have an effect in at least one tissue. On average, an inhibitor will alter the output activity, and potentially change the phenotypic response of the cell, in about 17 of the 84 tissues (~20%). 215
D.3 Diagram of the initial TCS-like Boolean network 217
D.4 Diagrams of each of 8 possible expression vectors for a hypothetical evolved network with three intermediate nodes. Beneath each subnetwork is the associated I/O Map, with different colors representing distinct maps. This example network thus presents 3 unique I/O Maps. 217
D.5 The flow chart for the evolution of the Boolean signaling networks. After a modification is made to the existing network, the intermediate nodes are variably expressed in every possible combination. We then count the number of unique I/O maps and keep the modification if it increases the number of unique maps. Otherwise the modification is removed and the process is started again. 218

List of Tables

B.1 Enzyme classification numbers used to search UniProt 164

B.2 The numbers and ratios of serine/threonine kinases and phosphatases from UniProt used in Figure 3.1 of Chapter 3. 164

B.3 The numbers and ratios of tyrosine kinases and phosphatases from UniProt used in Figure 3.1 of Chapter 3. 164

B.4 The numbers and ratios of the total number of phosphatases and substrates from UniProt used in Figure 3.1 of Chapter 3. 165

C.1 Table of coefficients for the linear fit of the log-log transformed K_A/K_S scatter plot from the interfacial residues of "K" domains' multiple sequence alignments 194

C.2 Table of coefficients for the quadratic fit of the log-log transformed K_A/K_S scatter plot from the interfacial residues of "K" domains' multiple sequence alignments . . 194

C.3 Table of coefficients for the linear fit of the log-log transformed K_A/K_S scatter plot from the noninterfacial residues of the "K" domains' multiple sequence alignments. 196

C.4 Table of coefficients for the quadratic fit of the log-log transformed K_A/K_S scatter plot from the noninterfacial residues of the "K" domains' multiple sequence alignments. 196

C.5 Table of coefficients for the linear model of the log-log transformed K_A/K_S scatter plot from the interfacial and noninterfacial residues of "K" domains' multiple sequence alignments. In this model, the parameter $Z = 1$ when K_S is from the "K" interface dataset and $Z = 0$ when K_S is from the "K" noninterface dataset. 197

C.6 Table of coefficients for the linear fit of the log-log transformed K_A/K_S scatter plot from the the average of 10 random subsets from the noninterfacial residues of the "K" domains' multiple sequence alignments. 197
C.7 Table of coefficients for the quadratic fit of the log-log transformed K_A/K_S scatter plot from the average of 10 random subsets from the noninterfacial residues of the "K" domains' multiple sequence alignments. 198
C.8 Table of coefficients for the linear model of the log-log transformed K_A/K_S scatter plot from the interfacial and the averages of 10 noninterfacial residues of "K" domains' multiple sequence alignments 199
C.9 Table of coefficients for the linear fit of the log-log transformed K_A/K_S scatter plot from the average of 10 random subsets from the noninterfacial surface residues of the "K" domains' multiple sequence alignments. 199
C.10 Table of coefficients for the quadratic fit of the log-log transformed K_A/K_S scatter plot from the average of 10 random subsets from the noninterfacial surface residues of the "K" domains' multiple sequence alignments. 200
C.11 Table of coefficients for the linear model of the log-log transformed K_A/K_S scatter plot from the interfacial and the averages of 10 noninterfacial surface residues of "K" domains' multiple sequence alignments 200
C.12 Table of coefficients for the linear fit of the log-log transformed K_A/K_S scatter plot from the interfacial surface residues of the "K" domains' multiple sequence alignments against HK853. 202
C.13 Table of coefficients for the quadratic fit of the log-log transformed K_A/K_S scatter plot from the interfacial surface residues of the "K" domains' multiple sequence alignments against HK853. 202
C.14 Table of coefficients for the linear fit of the log-log transformed K_A/K_S scatter plot from the noninterfacial surface residues of the "K" domains' multiple sequence alignments against HK853. 202

C.15 Table of coefficients for the quadratic fit of the log-log transformed K_A/K_S scatter plot from the interfacial surface residues of the "K" domains' multiple sequence alignments against HK853. 202
C.16 Table of coefficients for the linear model of the log-log transformed K_A/K_S scatter plot from the interfacial and the noninterfacial surface residues of "K" domains' multiple sequence alignments against HK853 203
C.17 Table of coefficients for the linear fit of the log-log transformed K_A/K_S scatter plot from the interfacial residues of the RRs' multiple sequence alignments. 203
C.18 Table of coefficients for the quadratic fit of the log-log transformed K_A/K_S scatter plot from the interfacial residues of the RRs' multiple sequence alignments. 203
C.19 Table of coefficients for the linear fit of the log-log transformed K_A/K_S scatter plot from the noninterfacial residues of the RRs' multiple sequence alignments. 204
C.20 Table of coefficients for the linear model of the log-log transformed K_A/K_S scatter plot from the interfacial and noninterfacial residues of RRs' multiple sequence alignments 204
C.21 Table of coefficients for the linear fit of the log-log transformed K_A/K_S scatter plot from 10 random subsets from the noninterfacial residues of the RRs' multiple sequence alignments. 205
C.22 Table of coefficients for the linear model of the log-log transformed K_A/K_S scatter plot from the interfacial and 10 random subsets from the noninterfacial residues of RRs' multiple sequence alignments 205
C.23 Table of coefficients for the linear fit of the log-log transformed K_A/K_S scatter plot from 10 random subsets from the noninterfacial surface residues of the RRs' multiple sequence alignments. 205
C.24 Table of coefficients for the linear model of the log-log transformed K_A/K_S scatter plot from the interfacial and 10 random subsets from the noninterfacial surface residues of RRs' multiple sequence alignments 206

D.1 Example I/O Map 217
D.2 The list of 29 canonical pathways in the KEGG Pathways database that were compiled to create the complete signaling network 221

Chapter 1

Introduction

Cells use a network of interacting proteins and other macromolecules to transmit information about their environment and respond to stimuli. These networks utilize a variety of information transfer mechanisms such as the diffusion of molecules, protein-protein interactions and covalent modifications. The post-translational covalent modification of proteins may include actions such as attaching a functional group (e.g., phosphorylation, acetylation, or glycosylation), changing the chemical nature of an amino acid (e.g., citrullination) or making structural changes (e.g., adding a disulfide bridge or removing a peptide chain) (Fig. 1.1*A*). These modifications alter the functionality of the protein. For example, the protein kinase Akt is catalytically inactive when dephosphorylated; phosphorylation at multiple sites is required in order for Akt to act as a kinase [1]. The reversible modification of proteins is used to transmit information about the cell's environment to the nucleus or other terminal location to bring about an appropriate phenotypic response (e.g., differentiation, proliferation, or apoptosis) (Fig. 1.1*B*).

Bacterial Two-Component Signaling (TCS) networks are comprised of many insulated signaling pathways. Typically these pathways include a membrane bound sensor histidine kinase (HK), which autophosphorylates in response to input. The HK binds and transfers its phosphoryl group to a response regulator (RR), which is usually a transcription regulator. HKs often are specific for a single RR, although a few, such as the phosphate sensor, PhoR, have been demonstrated to activate

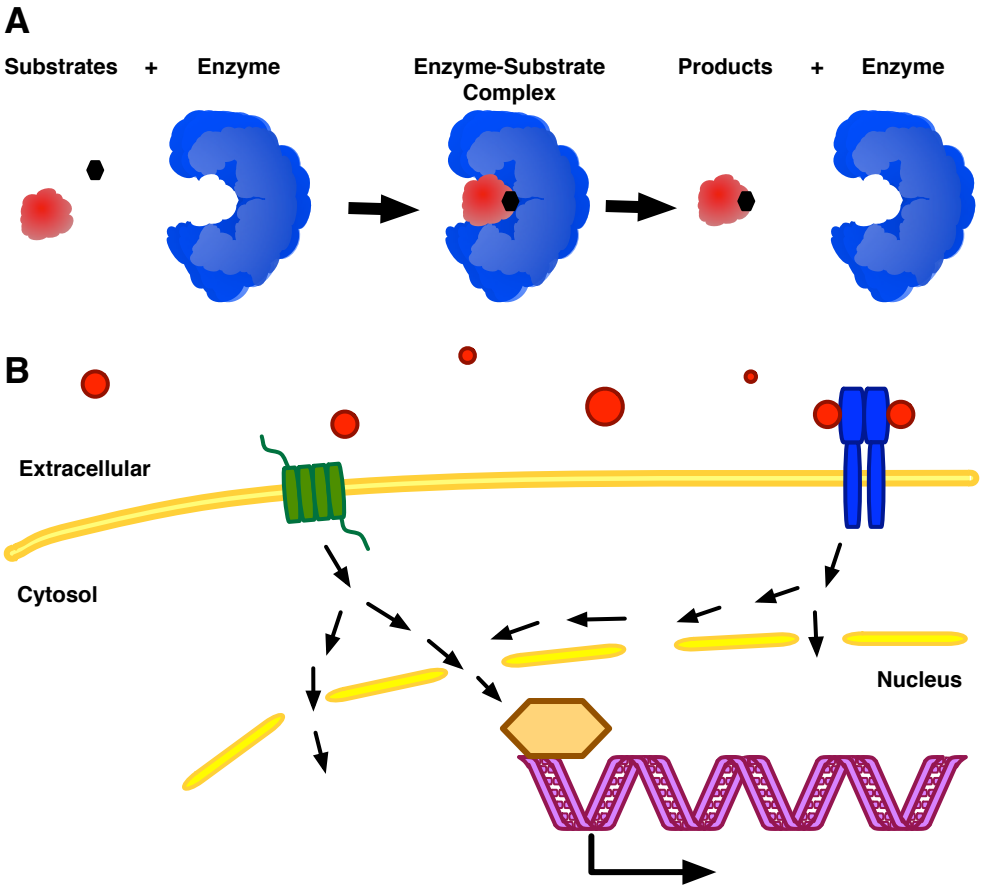

Figure 1.1: (*A*) An enzyme (blue) binds to and covalently modifies a substrate (red). After the reaction, the substrate keeps the modification, altering its functionality, and the enzyme is reset to its original state, ready to modify another substrate. (*B*) Cells use various receptors to intercept and transmit information from extracellular inputs through the membrane. The signal is then propagated through a network of interacting proteins, resulting in a phenotypic response to the stimuli (e.g., initializing the transcription of a gene or metabolizing a sugar molecule).

two or three RRs [2–9] (Fig. 1.2*A*). Metazoan signaling networks traditionally are considered to be comprised of many canonical pathways, each of which are activated by a different extracellular signaling molecule (e.g., epidermal growth factor (EGF), Wnt, or the insulin-like growth factor-I (IGF-I) signaling pathway) (Fig. 1.2*B*). These canonical pathways are generally thought to be unique entities in the cell, each using signaling cascades and other machinery in different ways in order to propagate information and respond to input. However, there have been many studies characterizing a phenomenon called network crosstalk, in which the activity of one pathway may affect the activity of other pathways [10–18]. In fact, Kirouac et al., have recently shown that none of the human canonical pathways meet even the simplest requirements to be considered a distinct entity within the cell. Many of the signaling interactions are revealed to actually be contributing to network crosstalk as they are used by multiple canonical pathways [19]. Although it is now commonly accepted that crosstalk is widespread in metazoan signaling networks, we do not yet have a clear conceptual understanding of why these signaling networks exhibit such a high degree of crosstalk and how the highly interconnected architecture may affect network behaviors and cellular responses.

In this work, we characterize the effects of crosstalk and competition between substrates for various signaling enzymes on the general behaviors and the evolution of signaling networks. In metazoans, these networks are comprised of many simple futile cycles in which one enzyme (e.g., a kinase) modifies a single substrate and another enzyme (e.g., a phosphatase) removes the modification. However, enzymes such as kinases and phosphatases often act upon a large number of different targets [22–24]. Competition between different substrates for the same kinase has been shown to greatly influence substrate phosphorylation [25].

Goldbeter and Koshland first characterized the simple motif of a kinase and a phosphatase regulating the phosphorylation of a substrate over 30 years ago, expressing the fraction of phosphorylated substrate as a function of the degrees of saturation and the ratio of the maximum velocities of the enzymes [26]. They found that saturation of the enzymes causes substrate phosphorylation to respond in a switch-like manner to the ratio of the maximum velocities of the enzymes, a behavior

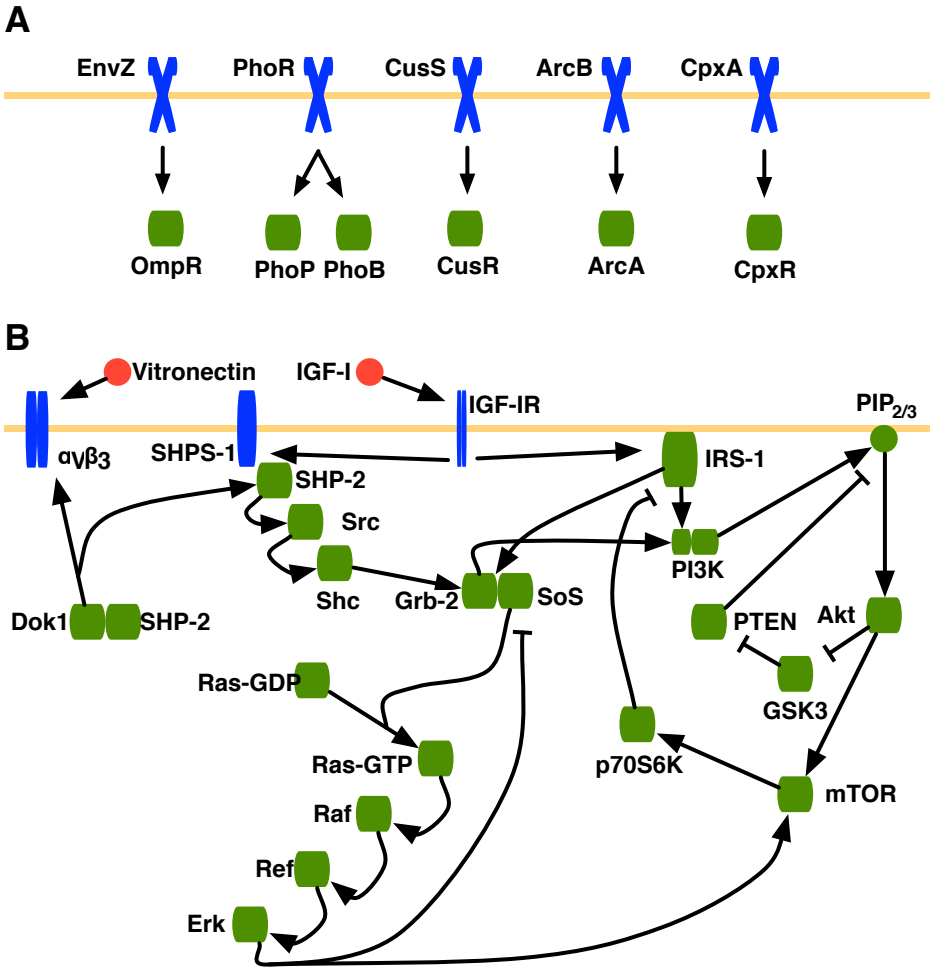

Figure 1.2: (*A*) Diagram of 5 TCS pathways found in bacteria. [20]. (*B*) Diagram of interactions between signaling molecules in the canonical IGF-I signaling pathway [18, 21].

referred to as 0^{th}-order ultrasensitivity. We build upon the simple futile cycle to include multiple substrates for either enzyme in order to investigate the effects of competition on substrate phosphorylation (see Chapter 2). The saturation of a shared enzyme becomes cumulative: if a single substrate saturates the enzyme, then it is saturated for every other substrate. Additionally, a set of substrates may collectively saturate the enzyme, even if none of the substrates would do so individually. A consequence of this collective saturation is that ultrasensitivity becomes transitive: the saturation of an enzyme by a subset of its targets causes every substrate to respond ultrasensitively to the enzyme, regardless of its individual saturation. Consistent with this is the demonstration of the ability of phosphatases to couple signaling responses on their own. Substrates of a shared phosphatase are able to respond to signals from kinases that do not directly act upon them. This is because an increase in phosphorylation levels in competing substrates decreases the concentration of phosphatase available for other substrates, leading to an increase in substrate phosphorylation. This indicates phosphatase-mediated crosstalk may be widespread in signaling networks, given the relative promiscuity of phosphatases when compared to kinases.

We find that the phosphoprotein/phosphatase ratio to be very large, with phosphatases having to act upon 30 substrates on average. Even if the substrates do not respond ultrasensitively to the phosphatase on their own, it is likely that they would collectively saturate the shared phosphatase. As such, the phosphorylation of any of the competing substrates would then monopolize the phosphatase, indirectly increasing the phosphorylation of the competing substrates. These findings suggest a potential paradox: each phosphatase must act on a large number of targets, yet the cell must avoid rampant phosphatase-mediated crosstalk between pathways.

We use mathematical models to explore different mechanisms by which phosphatases could provide insulation and identify trade-offs for each mechanism (see Chapter 2). An unsaturated phosphatase would provide insulation between its substrates. However, assuming biologically relevant concentrations of phosphoproteins, an unsaturated phosphatase would have to be an extremely inefficient enzyme. This would allow each of its substrates to respond to very little signal from its respective kinase. The synthesis and degradation of substrates could also act as a means

to dephosphorylate a pool of protein without needing a specific phosphatase. This would reduce the required number of targets per phosphatase and may decrease the potential for phosphatase-mediated crosstalk. In order for degradation to have a significant effect on substrate phosphorylation, the proteins would have to be highly unstable with half-lives on the order of hours. The rapid synthesis and degradation of substrates would thus be costly for the cell to maintain. Finally, we investigate whether the regulatory adaptor subunits may provide not only specificity but insulation between substrates. Phosphatases that have hundreds of documented substrates, such as PP1 and PP2A, use a set of regulatory adaptor subunits to target the catalytic core to specific substrates. These adaptor subunits can provide insulation between different sets of substrates dependent upon the manner of adaptor binding: the subunits must be able to bind their substrates independent of binding the catalytic core to provide such insulation. However, the phosphorylation of any substrate targeted by a specific regulatory adaptor subunit would still affect the phosphorylation of other targeted substrates.

These network behaviors, though, are based upon futile cycles using a pair of monofunctional enzymes. Unlike the majority of kinases in metazoans, HKs are bifunctional and are able to both phosphorylate and dephosphorylate the RRs [2, 3]. In contrast to metazoan signaling networks, TCS pathways exhibit very little crosstalk: HKs usually act upon a single RR [4–9]. Experimental studies have shown that HKs demonstrate a strong 'kinetic preference' for their cognate RRs [8, 9, 27–31]. However, it remains unclear exactly what mechanisms prevent TCS networks from evolving crosstalk.

In order to investigate the effects of competition and crosstalk on the activity of TCS pathways, we extend a well-studied and validated mathematical model of bifunctional HKs to include multiple interacting HKs and RRs (see Chapter 4) [32, 33]. Because HKs are both the kinase and the phosphatase, competition between multiple RRs decreases the phosphorylation of each of the substrates. These findings suggest a barrier in the evolution of new TCS pathways: immediately after duplicating an HK-RR pair, the new pathway would be able to interact with the parent pair, introducing a natural source of crosstalk. Using a coarse-grained model of the postduplication

divergence of these pathways, we characterize a set of 'near-neutral' evolutionary trajectories in which the new pathway may establish its own functionality while minimizing its impact on existing signaling pathways. All of these trajectories share a common order of events: the crosstalk between HK-RR pairs is removed prior to establishing the new input/output functionality of the new pathway. To test this prediction, we separately aligned and analyzed the PAS sensor and HK domain sequences from fully sequenced bacterial genomes. The interaction interface of the HK domains is revealed to experience strong positive selection immediately after duplication, likely due to the pressure to insulate the pathways. The PAS sensor domains often evolve through 'domain swapping', providing the HK with a new input functionality. Our sequence analyses indicate that the PAS sensor domains are generally swapped after the HK interfaces have been allowed to evolve RR specificity, as the near-neutral evolutionary trajectories predict.

Metazoan signaling networks do not have restraints on the evolution of crosstalk like those found in bacterial TCS networks. Many of the signaling proteins are shared between different canonical pathways, resulting in a network with a high degree of crosstalk [19]. However, the lack of restraints on crosstalk does not explain why metazoans have evolved so much interconnectivity. We demonstrate how crosstalk has evolved to allow for the evolution of various cell types that have the potential to respond differently to the same external stimuli through the differential expression of various subsets of signaling proteins.

Using a Boolean model of a TCS-like network, we find that cells with two receptor pathways could only demonstrate up to 4 unique responses to inputs, dependent upon the expression of either of the pathways (see Chapter 4). In order for TCS-like networks to allow cells to respond differently to the same stimuli, the cells would need to express a unique receptor pathway for every desired cellular response. Adding crosstalk to these networks through the inclusion of just sixteen intermediate nodes, however, can provide these cells with hundreds of different responses. These responses arise from differences in the expression of the intermediate nodes: the removal of a set of nodes may change the expressed topology of the network, altering how a receptor is able to propagate a signal from its environment to the network outputs. This behavior arises from

'compact coding': complex networks can achieve a larger diversity of phenotypic responses while maintaining a relatively low number of unique signaling molecules due to the extent of crosstalk.

The extensive crosstalk found in the human signaling network should make our cells more evolvable than cells possessing a TCS-like network, allowing cells to more easily adapt and provide new phenotypic responses through changes in the expression of signaling proteins. We find that the structure of the expressed signaling network changes in different tissues, just as the expressed topology did in our Boolean models. Despite the differences in structure, the expressed signaling networks retain a high degree of input/output connectivity when compared to networks with nodes expressed at random. Because of this connectivity, the expressed networks demonstrate a high level of diversity in responses to stimuli, dependent upon the specific expression of signaling proteins. These results have implications for drug studies: the introduction of an inhibitor specific to a signaling protein may not have an effect in every tissue, either because the protein is not expressed in every tissue or because of the presence of redundant pathways within the network. In tissues susceptible to the drug, inhibition of the protein results in different changes to the responses of these cells to stimulus.

Ultimately, this work demonstrates that the dynamics of enzymes and the architecture of the signaling networks can greatly influence the network behaviors and response of the cell to stimuli. Our results have implications for how we understand the role and evolution of crosstalk in different organisms. A major goal of systems biology is the construction of formal models of cellular regulatory systems. However, care must be taken to consider the cellular context in interpreting results of in vivo studies of network activity and structure.

Chapter 2

Crosstalk and Competition in Signaling Networks

2.1 Introduction

Signal propagation through a network of interacting proteins is central to a cell's ability to process and respond to stimuli. In most cases, these interactions involve an enzyme (e.g., a kinase) that covalently modifies a substrate and changes its functionality (i.e., activates/deactivates it as an enzyme, or causes translocation to a different compartment). To regulate the signal, another enzyme (e.g., a phosphatase) reverses the modification, restoring the original functionality of the substrate in question. The net activity of these enzymes alters the functional state of the proteins in the network in response to inputs, and the overall state of the network ultimately determines the cellular response.

Intracellular signaling networks are extremely complex in metazoans, which makes it difficult to understand their behavior [34, 35]. A major source of this complexity is network crosstalk, i.e., the sharing of input signals between multiple canonical pathways [12–14, 16, 17]. For example, kinases can often transmit signals to a large number of different targets: Akt can act on at least 18 substrates, and the receptor tyrosine kinases in the EGF/ErbB family can interact with >20

substrates [22, 23]. Because eukaryotic genomes contain fewer distinct phosphatases than distinct kinases, phosphatases are generally considered more promiscuous, and even with adaptor proteins targeting their activity, they often act on multiple substrates [24]. Although it is clear that crosstalk is widespread in mammalian signaling networks, we currently do not have a clear conceptual picture of how this highly interconnected architecture might influence the response of a network to incoming signals.

In this work, we seek to understand how the competition and promiscuity induced by crosstalk ultimately influence network behavior. In classic crosstalk, a kinase is shared between two pathways and can transfer signals from one pathway to another [12, 14, 16, 36]; for instance, mitogen-activated protein kinase (MAPK) networks often use the same enzymes in multiple cascades [37]. Most previous computational studies on this subject have focused on characterizing the spatial or temporal mechanisms for the insulation of MAPK signaling cascades despite the potential for crosstalk [38–40]. It has been demonstrated, however, that competition among targets of the same kinase can have profound effects on substrate phosphorylation [25]. Here, we extend these previous findings to characterize in detail how crosstalk can actively couple the response of multiple proteins to incoming signals. We developed models that consider a set of general motifs, with the goal of understanding how features such as substrate saturation and phosphatase architecture can influence substrate response.

Our models build off a simple futile cycle in which one enzyme modifies a single substrate and another enzyme removes the modification, which we represent as a kinase and phosphatase pair interacting with a target protein (see Fig. 2.1*A*). As first shown by Goldbeter and Koshland [26] over 30 years ago, the fraction of modified substrate for this cycle can be expressed as a function of three parameters:

$$K_K = \frac{K_{m,K}}{[S]_0}, \qquad K_P = \frac{K_{m,P}}{[S]_0}, \qquad r = \frac{V_{max,K}}{V_{max,P}} \tag{2.1}$$

where $[S]_0$ is the total amount of substrate, $K_{m,K}$ and $K_{m,P}$ are the Michaelis constants for the two enzymes, K_K and K_P represent the inverse of the degree of saturation of the enzymes, and r is the ratio of their maximum velocities. Detailed definitions of these constants in terms of underlying

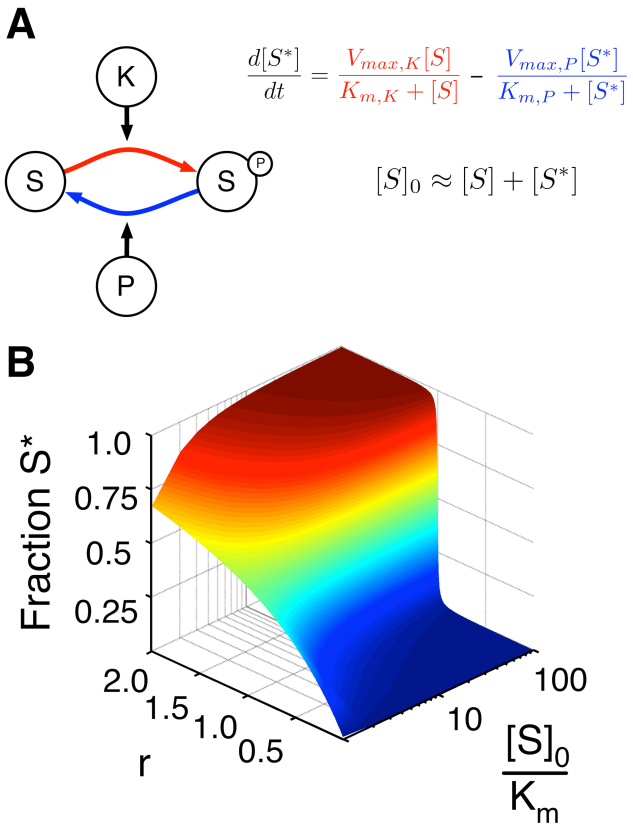

Figure 2.1: The Goldbeter-Koshland loop. (*A*) A pair of enzymes(say, a kinase *K* and a phosphatase *P*) acts on a single substrate. The associated equations show the change in S^* concentration as the difference between the production of S^* by the kinase (in *red*) and the production of *S* by the phosphatase (in *blue*). Here we assume that the concentration of free *S* and S^* is far greater than the concentrations of bound *S* in either form, which is necessary to obtain the standard Michaelis-Menten forms for the enzymatic reaction velocities [26]. (*B*) The fraction of phosphorylated *S* (z axis) is a function of r and $[S]_0$. The total concentration of $[S]$ is normalized by its K_m (which is identical for both the kinase and phosphatase) and is plotted on a log scale.

rates of the enzymatic reactions can be found in the context of Eq. 2 below. One can easily solve the underlying system of differential equations (see Fig. 2.1*A*) at steady state, providing a relationship between overall substrate phosphorylation and the parameters listed in Eq. 2.1 (see Eq. 3 below, with $\alpha_{K,1} = \alpha_{P,1} = 1$). Because protein levels tend to change slowly [41], we expect that saturation (and thus K_K and K_P) will remain constant on short timescales during the response to signal. On the other hand, r changes with the concentration of active kinase and phosphatase. Incoming signals generally modulate active K or P concentration, thus making r. However, when the substrate saturates both enzymes, the loop displays a switch-like behavior in r, referred to as 0^{th}-order ultrasensitivity (Fig. 2.1*B*). In this case, at values of $r < 1$ the fraction of phosphorylated substrate is very low, and at $r > 1$ the system switches to a highly phosphorylated state [26]. The ultrasensitive response of a substrate at saturating concentrations has been observed experimentally in a number of systems [25, 42–46].

We expanded this model to include competing substrates at either or both enzymes to characterize the influence of multiple targets on signaling (Fig. 1.2, *A*-*C*). All three of the motifs we consider are found in well-known signaling systems, such as the Fus3/Cdk1 network in yeast and other eukaryotes (Fig. 1.2*D*). We found that shared signaling enzymes can couple the responses of different substrates. For instances, when there is more than one substrate of the same kinase and phosphatase (see Fig. 1.2*A*), if one substrate is at sufficient concentration to elicit an ultrasensitive response, then all substrates that share the pair of enzymes in the cycle will exhibit ultrasensitivity without necessarily saturating the enzymes themselves. We have shown that in systems in which two substrates share a phosphatase (see Fig. 1.2*C*), one substrates saturating the phosphatase can cause the other substrate to ultrasensitively respond to signals from the first kinase. This indicates a novel potential for phosphatases to be involved in network crosstalk.

Kinases are becoming increasingly popular drug targets in the treatment of cancer and other diseases [53]. We considered how such inhibitors might influence the behavior of these various crosstalk architectures, and found that these inhibitors can have important consequences that would be difficult to predict in the absence of a detailed understanding of network topology and enzyme

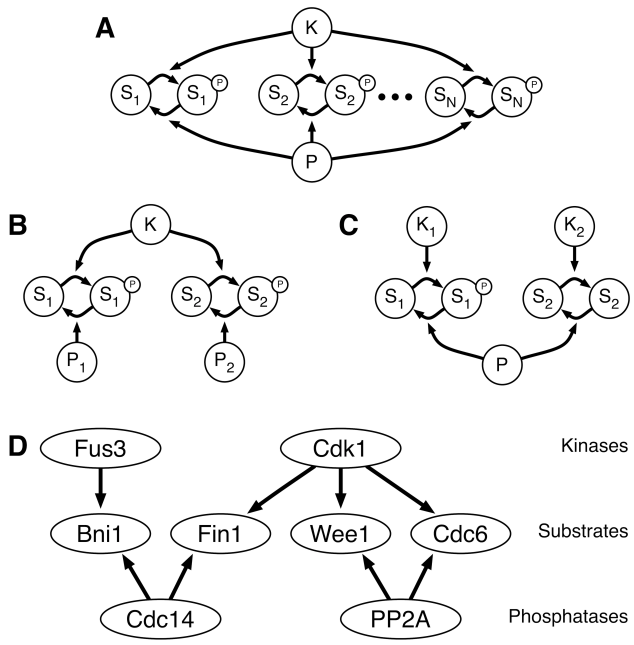

Figure 2.2: Crosstalk schematic. (*A*) A pair of enzymes (a kinase K and phosphatase P) acting on N substrates; we therm this the 1K1P loop. (*B*) A kinase that has two substrates, each with its own independent phosphatase (P_1 and P_2); we term this the 1K2P loop. (*C*) Two independent kinases (K_1 and K_2) acting on two substrates that share a single phosphatase P; we term this the 2K1P loop. (*D*) A section of the yeast Cdk1 signaling network, including each of these three motifs [25, 47–52]. Although the interactions shown are specific to yeast, there are human homologs for each of the proteins listed. The full network in this case contains a number of downstream feedback mechanisms that are omitted for clarity. These mechanisms may be abrogated by mutations so that the local influence of competition can be studied experimentally [25]. The competition between Wee1 and Cdc6 is an example of the 1K1P loop, whereas Wee1 and Fin1 form a 1K2P loop, and Fin1 and Bni1 form a 2K1P loop.

saturation.

Overall, our work demonstrates that enzymes with multiple targets can couple signal responses, and that systems considered in a cellular context may exhibit behaviors vastly different from those considered in isolated models. These results have implications for how we understand the role of crosstalk in signaling and how we can potentially control the propagation of the effects of enzymatic inhibitors through highly connected networks.

2.2 Materials and Methods

The behaviors of each model are described by sets of ordinary differential equations (ODEs), which are written explicitly for each system in section 1 of Appendix A. The systems of ODEs were numerically integrated using the CVODE package from SUNDIALS [54]. We employed the dense linear solver with the backward differentiation formula and the Newton iteration methodology available in that package for all of the dynamics discussed in this work. The values of the parameters used in each case are included in Appendix A.

Steady-state measurements were obtained by allowing the system to run until the level of each species of the system stabilized. The actual times at which the measurements were made were chosen heuristically by visual inspection of the trajectories themselves. The surfaces obtained in Figures 1.3 and 1.4 were confirmed analytically by solving for S_1^* in the same manner as described by Goldbeter and Koshland [26]. The analytical results are derived in sections 2-4 of Appendix A.

2.3 Results

2.3.1 1-Kinase/1-Phosphatase loop with two substrates

We first considered a signaling motif in which a kinase (K) and phosphatase (P) act on multiple substrates, which we term the 1-Kinase/1-Phosphatase (1K1P) loop. An example of this can be found in yeast, where the proteins Wee1 and CDC6 compete for both the kinase CDK1 and

phosphatase PP2A (Fig. 2.2*D*). In the simplest case, we included two substrates of the kinase and phosphatase, S_1 and S_2, each of which can exist in an unphosphorylated and phosphorylated (e.g., S_1^*) form (see Fig. 2.2*A*, $N = 1$). The set of enzymatic reactions is as follows:

$$\begin{aligned}
S_1 + K &\underset{k_{-,K,1}}{\overset{k_{+,K,1}}{\rightleftharpoons}} KS_1 \xrightarrow{k_{cat,K,1}} S_1^* + K \\
S_2 + K &\underset{k_{-,K,2}}{\overset{k_{+,K,2}}{\rightleftharpoons}} KS_2 \xrightarrow{k_{cat,K,2}} S_2^* + K \\
S_1^* + P &\underset{k_{-,P,1}}{\overset{k_{+,P,1}}{\rightleftharpoons}} PS_1^* \xrightarrow{k_{cat,P,1}} S_1 + P \\
S_2^* + P &\underset{k_{-,P,2}}{\overset{k_{+,P,2}}{\rightleftharpoons}} PS_2^* \xrightarrow{k_{cat,P,2}} S_2 + P
\end{aligned} \tag{2.2}$$

Each of the above reactions involves three elementary rates: the rate of complex formation (k_+), the rate of complex dissociation (k_-), and the enzyme catalytic rate (k_{cat}). From these rates we can obtain the Michaelis constant for both enzymes: $K_{m,K,i} = (k_{-,K,i} + k_{cat,K,i})/k_{+,K,i}$ and $K_{m,P,i} = (k_{-,P,i} + k_{cat,P,i})/k_{+,P,i}$. Additionally, we can define the maximum velocity of each enzymatic reaction as $V_{max,K,i} = [K]_0 k_{cat,K,i}$ and $V_{max,P,i} = [P]_0 k_{cat,P,i}$. Each kinase and phosphatase molecule can only bind and act on one substrate at any given moment, and as such, S_2 acts as a competitive inhibitor of the kinase and phosphatase reactions with S_1. This results in a set of inhibitory constants, $\alpha_{K,1} = 1 + [S_2]/K_{m,K,2}$ and $\alpha_{P,1} = 1 + [S_2^*]/K_{m,P,2}$, that capture the effects of S_2 on the S_1 kinase and phosphatase reactions, respectively. S_1 inhibition of the S_2 reactions generates similar constants, $\alpha_{K,2}$ and $\alpha_{P,2}$ (see Appendix A). The fact that multiple targets constitute competitive inhibitors of each other has been observed experimentally for both kinases and phosphatases [25, 55, 56]. These α terms are identical to what one would obtain for a generic competitive inhibitor, $\alpha = 1 + [I]/K_I$ [57]. Where the activity of a generic inhibitor against its target enzyme depends solely on its concentration, a competitive substrate will inhibit either the kinase or the phosphatase based on the concentrations of its unphosphorylated and phosphorylated forms, respectively. Because these concentrations are controlled by incoming signals, mutual inhibition has the potential to couple substrate responses.

The chemical reactions in Eq. 2.2 can be readily used to define a system of ODEs in which the binding, dissociation, and catalysis steps are treated explicitly (see Appendix A). We numerically integrated these equations and calculated the fraction $S_1^* \equiv [S_1^*]/[S_1]_0$ at steady state at various concentrations of S_2 for a case in which S_1 does not saturate the enzymes. In this work, we consider a case in which the saturation of all enzymes by any given substrate is equal; we leave the case of differential saturation among enzymes [37] to future studies. The response of the system is controlled by two r values, r_1 and r_2, which are the ratios of the maximum velocities of the enzymes with respect to either substrate. The results of these calculations are summarized in Fig. 3*A*. As expected, when there is no S_2 present to compete with S_1 displays an ultrasensitive response in r_1 in a fashion similar to the ultrasensitive response obtained by increasing S_1 concentration in Fig. 2.1*B*.

These findings can be understood by treating the 1K1P loop analytically. In the limit in which the total concentration of the substrates is much larger than the total concentration of either enzyme (i.e., $[S_1]_0 \approx [S_i] + [S_i^*]$), we can calculate the fraction S_1^* as:

$$S_1^* = \frac{(r_1-1) - (\alpha_{K,1}K_{K,1} + \alpha_{P,1}r_1K_{P,1}) + \sqrt{((r_1-1) - (\alpha_{K,1}K_{K,1} + \alpha_{P,1}r_1K_{P,1}))^2 + 4(r_1-1)\alpha_{P,1}r_1K_{P,1}}}{2(r_1-1)} \tag{2.3}$$

which is identical to the original result of Goldbeter and Koshland [26] except for the α inhibition terms (see Appendix A for details about the solution). Note that S_1^* depends on $[S_1]_0$ through the K terms as well as $[S_2]$ and $[S_2^*]$ through the α terms. The equation for S_2^* is identical to Eq. 2.3 with a change of indices. This result is a generalization of previous findings on multiple subtrates in a Goldbeter-Koshland loop, allowing for both kinase saturation and saturation of a shared phosphatase [25]. When $[S_1]_0 \ll K_m$, as in Fig. 2.3*A*, $\alpha_{K,2} \approx 1$ and $\alpha_{P,2} \approx 1$. In this case, S_2 will behave as an isolated Goldbeter-Koshland loop and as such will display an ultrasensitive response in r_2 when $[S_2] \gg K_m$. Because invoming signals vary r by changing the relative concentrations of active enzymes, $r_1 \propto r_2$ (for purposes of display in Fig. 2.3*A*, we assumed $r_1 = r_2$). When $r_2 < 1$, S_2 will be largely unphosphorylated and will inhibit the kinase's action on S_1, causing S_1 to be

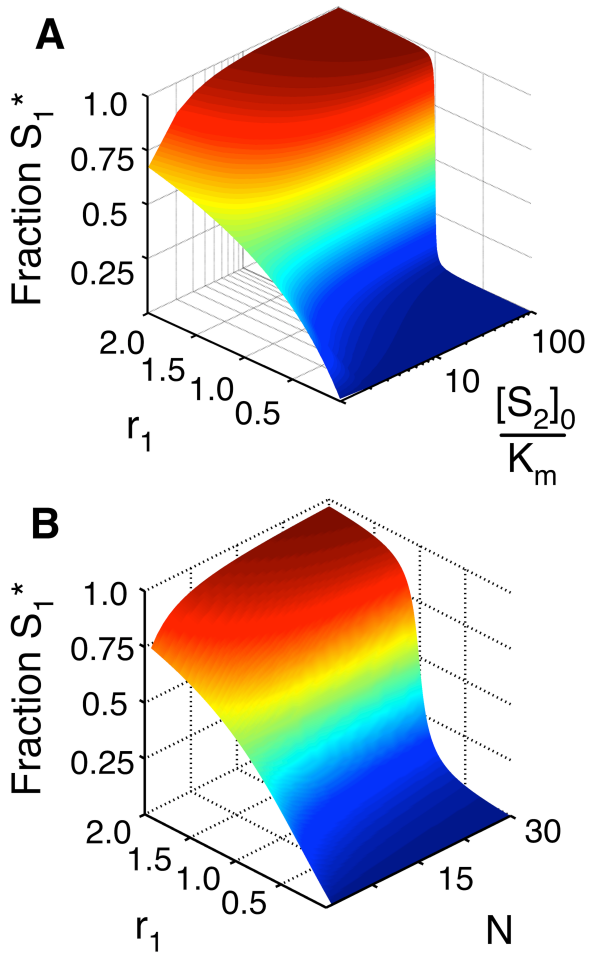

Figure 2.3: Results for the 1K1P loop. (*A*) The fraction of phosphorylated S_1 (z axis) as a function of r_1 and $[S_2]_0$. Note that for the purpose of display, we have set $r_1 = r_2$ in this case. The total concentration of $[S_2]$ is normalized by its K_m (which is identical for both the kinase and phosphatase) and is plotted on a log scale. (*B*) The fraction of phosphorylated S_1 as a function of r_1 and the number of additional substrates in the loop (N, see Fig. 2.2*A*). All substrates are below saturating concentrations ($[S_i]_0 = 0.1 \times K_m$). As in *A*, for the purpose of display, the r and K_m parameters have been set to be equal for all substrates. Note that in both panels *A* and *B*, the fraction S_1^* responses to r_1 with increasing ultrasensitivity as the total saturation of the enzymes (represented by $[S_2]_0/K_m$ or N, respectively) increases.

primarily in its unphosphorylated state. Similarly, when $r_2 > 1$, S_2 will be mostly phosphorylated and will inhibit the S_1 dephosphorylation reaction by saturating the phosphatase. In combination, this coupling transfers the ultrasensitive resposne of S_2 to the S_1 curve. We have proven mathematically that an increase in S_2 ultrasensitivity (i.e., increasing S_2 concentration) always increases the ultrasensitivity of the response of S_1 in r_2 regardless of the values of the kinetic parameters (see Appendix A). The general behavior observed in Fig. 2.3*A* is thus a qualitative feature of all 1K1P loops.

It has been shown experimentally that the competition between multiple phosphorylation sites on the protein Wee1 contributes to the ultrasensitivity of Wee1's response to incoming signals [25]. Although multisite phosphorylation can have a number of influences on such systems (e.g., by introducing thresholds or bistability [35, 58, 59], these findings are consistent with the predictions made by Eq. 2.3.

2.3.2 1-Kinase/1-Phosphatase loop with many substrates

We further developed the 1K1P loop to include $N > 2$ substrates of the kinase and phosphatase (see Fig. 2.2*A*). As described above, we numerically integrated the resulting ODEs and calculated the fraction S_1 at steady state in a case in which we include a varying number of substrates, each of which does not saturate the enzymes. The results of these calculations are summarized in Fig. 2.3*B*. As expected, S_1^* increases as a rectangular hyperbola in r_1 in the absence of other substrates. As new unsaturating substrates are added to the system, we see that S_1^* starts to show an ultrasensitive response in r_1, even though none of the substrates are at a concentration that would produce such a response on their own.

Once again, these results can be understood by treating the loop analytically. In this case, the collection of substrates act as competitive inhibitors of the S_1 loop. As such, the inhibitory constants must now account for all competing substrates and can be expressed as $\alpha_{K,1} = 1 + \sum_{i=2}^{N}[S_i]/K_{m,K,i}$ and $\alpha_{P,1} = 1 + \sum_{i=2}^{N}[S_i^*]/K_{m,P,i}$ (see Appendix A for the derivation). Considering the case in which $N > 2$ reveals that saturation of the enzymes can be the combined result of many

substrates, rather than one substrate saturating the enzymes on its own. When the kinase is saturated by any subset of its targets, S_1's kinase reaction is inhibited, and a similar inhibition occurs with the phosphatase. Thus, given enough substrates, the entire system can show ultrasensitivity in r_1 even when none of the substrates individually saturate the enzymes.

As mentioned in the Introduction, kinases often have multiple targets within cells; for instance, Cdk1 has hundreds of substrates in yeast [35, 60, 61], and the ErbB receptor tyrosine kinases in humans have between 20 and 40 potential targets. In the latter case, the K_D values measured by [62] indicate that the 1 μM K_m value used in generating Fig. 2.3 is a reasonable estimate. The collective-saturation mechanism described above may thus represent a common scenario for generating ultrasensitivity in substrate response.

2.3.3 1-Kinase/2-Phosphatase loop

Most of our empirical understanding of crosstalk comes from studies that focused on the motif of a kinase with more than one substrate [63]. Because the specific phosphatases that act on any given set of targets are often not known, it is not clear that all kinase crosstalk will follow the 1K1P pattern discussed above (Fig. 2.2*A*). For instance, Fin1 and Wee1 share the same kinase (Cdk1) but have separate phosphatases (Cdc14 and PP2A, respectively; Fig. 2.2*D*). Also, because kinases often have a very large number of targets, systems in which substrates share the same kinase but possess separate phosphatases may be widespread [22, 23, 60, 61]. As such, we considered the behavior of the 1-Kinase/2-Phosphatase (1K2P) loop as diagramed in Fig. 2.2*B*. In this case, because the phosphatases are independent, we can separate the r parameters (i.e., $r_2 \not\propto r_1$). At low substrate concentrations, S_1 responds hyperbolically in r_1 and is insensitive to r_2 (Fig. 2.4*A*). When $[S_2]_0 \gg K_m$ and $r_2 < 1$, S_1 phosphorylation is greatly reduced (Fig. 2.4*B*). In fact, one observes very little S_1 phosphorylation until $r_2 > 1$. In contrast to the 1K1P loop, the response of S_1 to r_2 thus exhibits a threshold: when $r_2 < 1$, S_1 essentially cannot respond to signals. At values of $r_2 > 1$, however, S_1 responds hyperbolically to both r_1 and r_2.

The fraction S_1^* for the 1K2P loop also follows Eq. 2.3, but with $\alpha_{P,1} = 1$ because the phos-

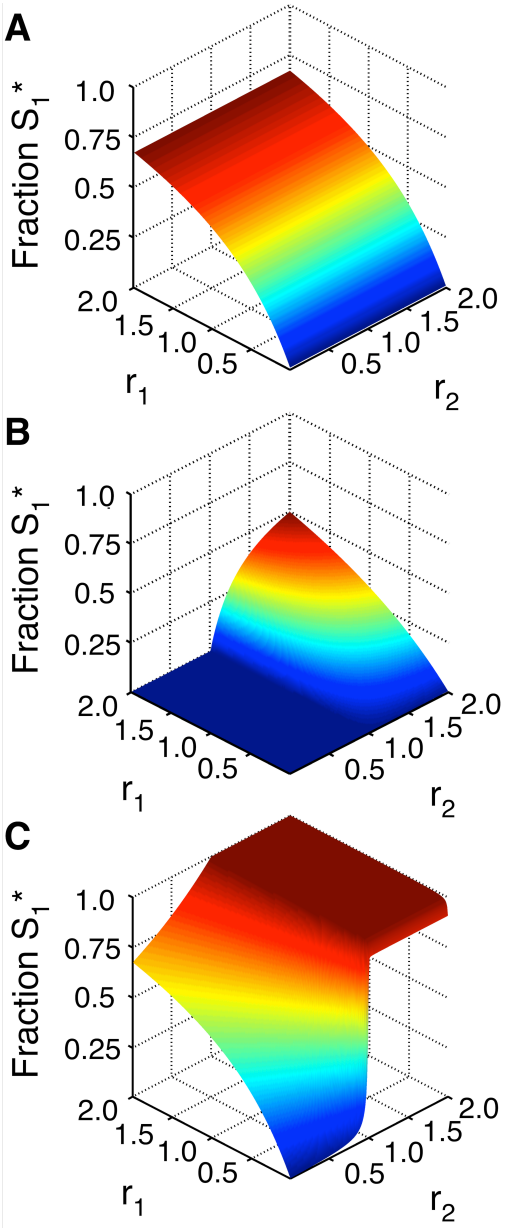

Figure 2.4: Influence of phosphatase architecture on network response. (*A*) The fraction of phosphorylated S_1 as a function of r_1 and r_2 when $[S_2]_0 \ll K_m$ for both the 1K2P and 2K1P loops. In this case, $[S_1]_0 = 0.1 \times K_m$. Note that r_2 has little effect on the response of the S_1 loop. (*B*) The fraction of phosphorylated S_1 as a function of r_1 and r_2 for a 1K2P loop with $[S_2]_0 = 20 \times K_m$. As in *A*, $S_1 = 0.1 \times K_m$. If S_2 saturates the enzymes, it becomes a gatekeeper, when $r_2 < 1$ (i.e., when the S_2 loop is switched to the unphosphorylated state), the S_1 loop essentially cannot respond to incoming signals. When $r_2 > 1$, however, S_1^* responds hyperbolically in both r_1 and r_2. (*C*) The fraction of phosphorylated S_1 as a function of r_1 and r_2 for a 2K1P loop. As in *B*, $[S_1]_0 = 0.1 \times K_m$. Saturating concentrations of S_2 generally increase phosphorylation in this case. Note that even when $r_1 \ll 1$, S_1 shows an ultrasensitive response to r_2 (and thus K_2) despite receiving only basal levels of signal from its own kinase. This indicates the potential for significant phosphatase crosstalk in signaling networks.

phatases are independent. The presence of S_2 in the system thus generally decreases the phosphorylation level of S_1 (compare Fig. 2.4*A* and *B*). The thresholding behavior seen in Fig. 2.4*B* occurs because the concentration of the inhibitor (i.e., unphosphorylated S_2) responds ultrasensitively to r_2. If $r_1 < 1$, the inhibitor concentration is high, and no phosphorylation of S_1 can take place. At $r_2 > 1$, the inhibitor is largely removed from the system, allowing S_1 to respond to incoming signals. However, it is only in the limit $r_2 \to \infty$ (i.e., $\alpha_{K,1} \to 1$) that S_1 will behave as an isolated futile cycle. As with the 1K1P loop, we have shown mathematically that addition of S_2 always decreases S_1^* regardless of the values of the parameters in the limit $S_1^* \ll K_m$ (see Appendix A). This indicates that the gatekeeper function played by S_2 is a robust feature of 1K2P loops.

Kim and Ferrell showed experimentally that adding Fin1 and Cdc6 to *Xenopus* cell extracts increases the active kinase concentration (i.e., r) required to induce a Wee1 response [25]. Although the experiment in this case involves both a 1K1P and a 1K2P loop (Fig. 2.2*D*), these findings are consistent with our prediction that competitive substrates tend to decrease the phosphorylation levels of other targets when the phosphatase is not shared.

2.3.4 2-Kinase/1-Phosphatase loop

The human genome encodes 150 catalytically active phosphatases and phosphatase domains, and almost 500 kinases [64, 65]. As such, phosphatases are generally considered promiscuous; although adaptor proteins help increase phosphatase specificity, these complexes still can target multiple substrates [24]. Because of this promiscuity, it is reasonable to imagine that motifs in which two substrates share a single phosphatase but are phosphorylated by independent kinases are relatively common arrangements in signaling networks. There are certainly examples of such situations: for instance, Fin1 and Bni1 in yeast share a phosphatase (Cdc14) but have different kinases (Cdk1 and Fus3, respectively; Fig 2.2*D*). We used the 2-Kinase/1-Phosphatase (2K1P) loop as modeled in Fig. 2.2*C* to characterize the behavior of such systems. As with the 1K2P loop, the distinct kinases in the 2K1P system allow the separation of r parameters so that $r_1 \not\propto r_2$.

At low substrate concentrations, this is essentially the case. As anticipated, S_1 responds hy-

perbolically in r_1 and is insensitive to r_2 (see Fig. 2.4*A*). The situation is very different when $[S_2]_0 \gg K_m$. We see the expected hyperbolic S_1 response in r_1 when r_2 is nearly zero (i.e., when the S_2 loop has not received and activation signal); however, as r_2 increases, the fraction of phosphorylated S_1 molecules increases until it reaches nearly one at $r_2 > 1$ (Fig. 2.4*C*). When r_1 is close to zero, S_1 responds ultrasensitively to r_2. This indicates that a signal that switches S_2 to its phosphorylated state can cause a similar switch in S_1 even if very little signal is received via K_1.

As with the 1K1P loop, this behavior can be explained in terms of the inhibition of one loop by another. In this case, the fraction of S_1^* can be defined as in Eq. 2.3 with $\alpha_{K,1} = 1$ to account for the independence of the kinases. Adding S_2 to the system thus generally increases phosphorylation of S_1 (compare Fig. 2.4*A* and *C*). Because phosphorylated S_2 acts as a phosphatase inhibitor, an incoming signal that increases r_2 to values greater than one introduces high concentrations of the inhibitor in a switch-like manner, inducing a response in S_1. We have shown mathematically that this increase in phosphorylation in response to S_2 competition will always occur regardless of parameters in the limit $S_1^* \ll K_m$ (see Appendix A).

2.3.5 Phosphatase tunneling

In the models described above, we focused on crosstalk occurring between substrates on the same level of signaling; the only relationship between the substrates is the shared enzymes. Signaling networks, however, often contain cascades in which a set of proteins activate each other in sequence [66]. Although the sharing of phosphatases between different levels of a cascade has been documented [13], the phosphatase architecture in these cases is often poorly understood. Indeed, anonymous and independent phosphatases are often added to mathematical models of MAPK cascades to fill in these gaps [44, 67–69]. Given this ambiguity, we constructed models of cascades in which each kinase has an independent phosphatase, in addition to a case in which a single phosphatase acts on all of the proteins in the cascade (Fig. 2.5*A* and *B*).

Each type of cascade was modeled with depth $N = 2$, 3, 4, or 5 substrates present in saturation ($10 \times K_m$) or unsaturating ($0.1 \times K_m$) concentrations. The input parameter r was defined as the ratio

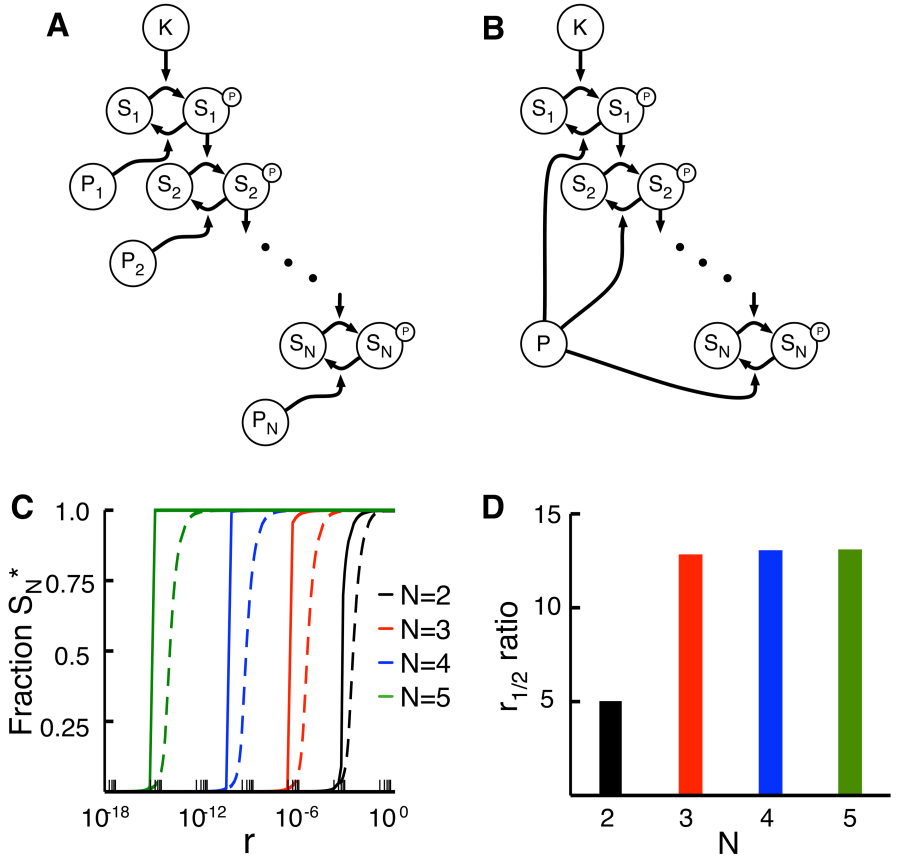

Figure 2.5: Influence of phosphatase tunneling on cascade signals. (*A*) A kinase with N members. The kinase K provides the input signal, and each substrate S_i acts as the kinase for substrate S_{i+1}. In this model, there are N independent phosphatases (P_i). This expands upon systems previously described by Goldbeter and Koshland [26]. (*B*) A kinase cascade similar to that in panel *A*, but with a single shared phosphatase P. (*C*) Fractional phosphorylation of the final substrate in the cascade as a function of r for cascades with two to five substrate. In this case, r is defined as the V_{max} of the input kinase (K in *A* and *B*) divided by the V_{max} of the phosphatase for the first substrate in the cascade (P_1 in *A* and P in *B*). The dashed lines represent cascades with N phosphatases and the solid lines represent cascades with a single shared phosphatase. Note that the responses of cascades become exponentially more sensitive to r with increasing depth N. Cascades with a single shared phosphatase are considerably more sensitive to r compared with those with independent phosphatases. (*D*) In this case, we define a parameter, $r_{1/2}$, as the value of r in panel *C* at which the response of a cascade is half-maximal. For any given number of substrates, N, the $r_{1/2}$ of the independent case divided by the $r_{1/2}$ of the shared case (i.e., the $r_{1/2}$ of the dashed curve in *C* divided by the $r_{1/2}$ for the solid curve). For $N = 2$, the independent case requires 5 times as much input signal to achieve a half-maximal response; for N =3, 4, and 5, the independent case requires 13 times as much input signal.

of the maximum velocities of the initial kinase (K) to the phosphatase acting on S_1 (P_1 or P for the independent and shared cases, respectively), and the models were analyzed for the fraction of the final substrate phosphorylated (S_N^*) at steady state.

For both classes of cascade, we found that the response of the final substrate becomes exponentially more sensitive to input signals with increasing cascade depth. The $N = 5$ case generally reaches its $r_{1/2}$ (the r-value at which half of S_N is phosphorylated) with 9 orders of magnitude less input than $N = 2$ (see Fig. 2.5*C*). This increase in sensitivity is an expected outcome of amplification in signaling cascades [26, 70]. Additionally, models with a single, shared phosphatase show a higher degree in input sensitivity in r compared with models with independent phosphatases, but only when the substrates are present in saturating concentrations. To quantify the changes in input sensitivity for saturating conditions, we took the ratio of the $r_{1/2}$-values for the two types of cascade at a given value of N (see Fig. 2.5*D*). In the most basic cascade, with $N = 2$, the $r_{1/2}$for the single phosphatase model is 5 times less than that for the multiple phosphatase model. This ratio increases and plateaus for cascades with depth $N \geq 3$; in these cases, the single phosphatase models require 13 times less signal. This occurs because the signal is able to tunnel through the shared phosphatase when the substrates are at saturating concentrations. Activation of the upstream kinases not only activates the rest of the cascade but also produces phosphorylated substrate molecules that act as phosphatase inhibitors. This reduces the effective concentration of free phosphatase available for downstream substrates, amplifying the apparent signal strength.

2.3.6 Kinase inhibitors

As mentioned above, there is a growing interest in developing small molecules that target and inhibit kinases as potential therapeutics for a variety of diseases [53]. It is unclear, however, what kind of effects these inhibitors will have in loops with significant kinase or phosphatase crosstalk; in these cases, kinase inhibitors not only influence their targets' activity but also the concentration of other inhibitors (namely, S_2 and S_2^*) in the system. We considered the impact of two separate types of inhibitors on the loops described above. Type 1 inhibitors, which are currently by far the

most commonly used in practice [53]., target the ATP-binding site of a specific kinase and disrupt its activity toward all of its targets. Type 2 inhibitors, on the other hand, target and disrupt a specific kinase-target interaction, leaving the kinase free to act on a subset of its other targets. Although the latter is not currently common, peptide inhibitors have been successfully used in this manner [56], and there is increasing interest in developing the capacity to inhibit specific protein-protein interactions within cells [71].

We modeled the potential effects of these inhibitors by including explicit inhibitor molecules in our loops, with I_1 and I_2 representing type 1 and type 2 inhibitors, respectively. We first considered a 1K1P loop with S_2 at saturating concentrations and in the active state ($r_1 = r_2 = 1.5$; see Fig. 2.2*A*). as one would expect, adding I_1 significantly decreases S_1^*, because a generic inhibitor for the kinase will clearly reduce overall phosphorylation of all targets (Fig. 2.6*A*). However, even an inhibitor that is specific to S_2 decreases the phosphorylation of S_1 (Fig. 2.6*A*). The specific inhibitor in this case decreases the concentration of S_2^*, reducing competition for the phosphatase and thus decreasing S_1^*. The effect of I_2 is not as dramatic as that of I_1 for the 1K1P loop, but this nonetheless represents a potentially unintended consequence of a (putatively) specific inhibitor.

In the 1K2P case, we find exactly the opposite behavior: whereas I_1 decreases S_1^* as expected, I_2 increases the phosphorylation of the first substrate (Fig. 2.6*B*). This is because the inhibitor reduces S_2 interactions with the kinase, alleviating competition. In this case, the response of the system is perhaps more intuitive: because S_2 is a competitive inhibitor of S_1 phosphorylation, inhibiting its phosphorylation in a specific way increases the capacity of S_1 to respond to signals.

In the 2K1P loop, if the two types of inhibitors are aimed at the second kinase (K_2), they have the same net effect. Because K_2 cannot act on S_1 in this model, there is no difference between an inhibitor that simply targets K_2 and one that specifically targets the $K_2 - S_2$ interaction. When the second loop is activated by a signal and the first loop is not, the K_2 completely abolishes S_1 phosphorylation (Fig. 2.6*C*).Although the source of this behavior is clear for Fig. 2.4*C*, the effect is nonetheless striking. In the absence of knowledge about the shared phosphatase (or the phenomenology of the 2K1P loop), a response like the one shown in Fig. 2.4*C* might lead to the

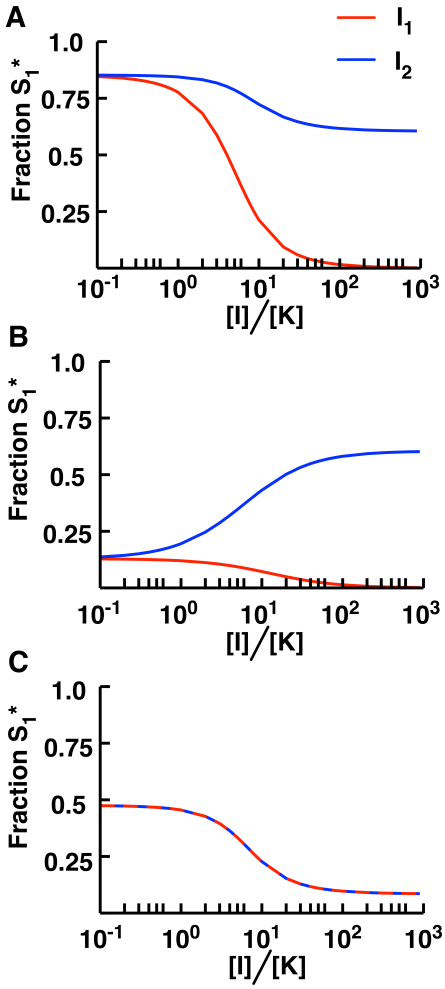

Figure 2.6: Effect of kinase inhibitors in the presence of crosstalk. (*A*) A 1K1P loop with two substrates in the presence of one of two kinase inhibitors: I_1, which prevents reactions with all targets of the kinase (*red*), or I_2, which specifically disrupts $K - S_2$ interactions (*blue*). We plot the fraction of phosphorylated S_1 against the ratio of $[I_1]$ or $[I_2]$ to $[K]$. In this case, $[S_1]_0 = 0.1 \times K_m$, $[S_2]_0 = 20 \times K_m$, and $r_1 = r_2 = 1.5$. Note that using either inhibitor causes a decrease in the fraction S_1^*, although the effect is less pronounced with the S_2-specific inhibitor. In the latter scenario, I_2 reduces the $[S_2^*]$, which is itself a phosphatase inhibitor for S_1^*. The net effect of I_2 is thus to decrease S_1 phosphorylation. (*B*) A 1K2P loop with the same kinase inhibitors as in panel *A*. The fraction of phosphorylated S_1 is plotted against the ratio of $[I_1]$ or $[I_2]$ to $[K]$. In this case, $[S_1]_0 = 0.1 \times K_m$, $[S_2]_0 = 20 \times K_m$, $r_1 = 0.5$, and $r_2 = 1.5$. Although the general inhibitor still reduces S_1^*, the specific inhibitor increases S_1^*. This is because decreasing the concentration of S_2^* reduces competition for the shared kinase. (*C*) A 2K1P loop in the presence of both I_1 and I_2. Note that because the kinases are independent in this case, the effects of both inhibitors are identical. The fraction of phosphorylated S_1 is plotted against the ratio of the concentrations of $[I_2]$ to $[K]$. In this case, $[S_1]_0 = 0.1 \times K_m$, $[S_2]_0 = 20 \times K_m$, $r_1 = 0.01$ and $r_2 = 1.5$. Both inhibitors decrease S_1^*, as the reduction in phosphorylated S_2 due to the inhibitors reduces S_2^*'s inhibition of the S_1 phosphatase reaction.

erroneous conclusion that K_2 acts directly on S_1, or that the inhibitor in this case is nonspecific.

2.4 Discussion

The 1K1P and 1K2P loops discussed above (Fig. 2.2*A* and *B*) represent two variations on the classic crosstalk motif, i.e., a kinase that has multiple downstream targets in different pathways. In the traditional view, the coupling between the substrates in these two loops is understood as simply arising from the fact that they will all respond to some of the same upstream signals [63]. Our work reveals that a shared enzyme not only modifies each target but also can strongly couple the response of one target to that of another through competitive inhibition at the shared enzyme. For instance, if the targets in question share the same kinase and phosphatase, we find that 0^{th}-order ultrasensitivity becomes transitive; all of the targets in this case will respond in a switch-like manner to incoming signals (Fig. 2.3*A*). We also find that in situations where there are a large number of substrates (Fig. 2.3*B*), the system can respond ultrasensitively even if none of the targets is at a high enough concentration to elicit such a response on its own. It has been shown that some kinases do in fact act on many targets (e.g., Akt, the EGF receptors, and Cdk1 [22, 23, 60, 61]), indicating that this collective saturation may represent a common mechanism for inducing ultrasensitivity without having to express any given protein target at saturating levels.

We find that the alternative variation on traditional kinase crosstalk, the 1K2P loop (Fig. 2.2*B*), displays a completely different set of behaviors from those observed when the phosphatase is shared. In this case, the saturating substrate acts as a type of gatekeeper for the other substrates in the loop. Below the signal threshold at which this saturating substrate switches into the phosphorylated state, other substrates will simply be unable to respond to incoming signals, whereas above this threshold the unsaturating targets will respond in a hyperbolic manner (Fig. 2.4*B*). Although direct experimental tests are currently lacking, our predictions for both 1K1P and 1K2P loops are consistent with available data [25]. Overall, these findings indicate that when a particular kinase has multiple targets in multiple pathways, it is difficult to reason in general about the behavior of

the system in the absence of detailed information regarding phosphatase architecture and relative saturation levels (Figs. 2.3 and 2.4).

To date, nearly all experimental characterizations of crosstalk have focused on kinases, and, to our knowledge, the potential for phosphatases to couple signaling responses on their own has not been previously considered [63]. Our analysis of the 2K1P loop (Fig. 2.2*C*) demonstrates that such coupling is readily achieved. Indeed, a shared phosphatase can elicit an ultrasensitive response of a target to signals from kinases that do not directly act on the target in question (Fig. 2.4*C*). Furthermore, phosphatase architecture plays a role in the sensitivity of a signaling cascade. We found that cascades in which every substrate shares a common phosphatase are more responsive to input signals than cascades with independent phosphatases when the substrates are at saturating levels. Given that phosphatases are generally considered more promiscuous than kinases, this indicates that phosphatase crosstalk may be widespread in biological networks. Because the specific phosphatases that act on many targets in signaling networks are often not known [67–69], it is currently unclear to what extent phosphatase crosstalk can influence global network behavior.

Given the widespread crosstalk present in mammalian signaling networks, our work highlights the inherent difficulty of predicting a priori the effects that kinase inhibitors will have on cells. These effects ultimately will depend not only on the kinase connectivity of the network but also on the degree of saturation in the targets and the phosphatase architecture. In many cases, both of these facts are unknown - even if the intracellular concentrations of the target proteins are known, the K_m-values for kinases and (especially) phosphatases are not known, and for many signaling pathways the relevant phosphatases have not yet been identified. Understanding these details will be a crucial component of any attempt to rationally design a kinase inhibition strategy that can elicit some desired effect on some set of targets without inducing unintended decreases (or increases) in the phosphorylation levels of other proteins in the network (Fig. 2.6).

Ultimately, our work indicates that studies on signaling and regulatory networks need to be increasingly mindful of the highly interconnected and interdependent structure of the networks themselves. This is especially true of phosphatases. To understand the real consequences of ram-

pant kinase crosstalk, we clearly must obtain more reliable information about which phosphatases act on which targets, what adaptor domains they employ, etc. The findings described above also highlight the fact that individual elements of signaling networks can exhibit responses that are sensitive to the context in which the element is found. Care must be taken to ensure that this dependence on network architecture informs our interpretation and understanding of how networks function and interact with each other.

Chapter 3

Phosphatase Specificity and Pathway Insulation in Signaling Networks

3.1 Introduction

Signaling networks allow cells to sense changes in their environment and respond adaptively. One of the most common "motifs" in eukaryotic signaling networks consists of a kinase that phosphorylates another protein in the network. Phosphorylation often alters the function of the target protein: for instance, that target might itself be a kinase that only becomes active when it is phosphorylated. A second enzyme, called a phosphatase, generally catalyzes the removal of the phosphoryl group. Although they are generally less well studied than kinases, phosphatases play a crucial role in controlling the phosphorylation levels of target proteins and thus the response of signaling networks to external stimuli [13, 26, 72].

Metazoan signaling networks are often very complex, exhibiting a high degree of "crosstalk" where many enzymes are shared between multiple pathways [12–14, 16, 17, 34, 35]. Crosstalk studies have generally concentrated on the interactions made by kinases, and the potential for phosphatases to contribute to signaling complexity has been largely overlooked [63]. For instance, in developing mathematical models of signaling networks, when a phosphatase has not yet been

identified for a particular phosphoprotein, an anonymous, independent, and often unsaturatable phosphatase is added to the model to fill in the gap [25, 44, 67, 69, 73–75]. This approach obviously ignores any contribution that phosphatase-mediated crosstalk might make to the behavior of the network.

We recently used a set of mathematical models to explore whether phosphatases acting on multiple substrates could impact signaling dynamics. We found that the responses of substrates to incoming signals can be strongly coupled if they share a phosphatase [72]. In particular, we considered a case where two different substrate proteins S_1 and S_2 in the network have two completely independent kinases but share a single phosphatase (i.e. the configuration diagrammed in Fig. 3.2*A* below). In this case, signals that activate only one of the kinases can cause both substrates to respond in a switch-like manner; this occurs because the phosphorylated substrate (say S_2^*) will act as a competitive inhibitor of the phosphatase, causing the other substrate to become active. This significantly increases S_1 phosphorylation even when the kinase specific to that substrate has very low activity.

While over 500 distinct kinases have been identified in the human genome, there are only a total of about 150 phosphatases [64, 65]. There are thus not even enough phosphatases to assign a unique one to each kinase, let alone o each unique substrate in the network, as has commonly been assumed in modeling studies [25, 44, 67, 69, 73–75]. In fact, it is well established that phosphatases are often inherently "promiscuous:" well-characterized phosphatases have been shown to act on tens if not hundreds of substrates [76–78]. This fact suggests a potential paradox: since, by virtue of their relatively small numbers, each phosphatase must act on a large number of targets, it is unclear how the cell avoids rampant phosphatase-mediated crosstalk between distinct parts of the network [72].

In this work, we used mathematical models to investigate a variety of mechanisms that cells could deploy to prevent shared phosphatases from resulting in unwanted crosstalk. These models focused on a simplified scenario in which substrates share a single phosphatase but are otherwise unrelated (e.g. Fig. 3.1*A*). While this ignores crosstalk at the kinase level [12, 14, 17, 34, 35, 44,

63, 72] and the fact that phosphatase activity is often itself regulated by the signaling network (24, 25), it allows us to isolate a particular source of crosstalk and characterize various mechanisms the cell might use to prevent it.

Since phosphatases can only couple substrate responses if they are saturated [72], one natural approach to limiting the impact of the phosphatase crosstalk would be to evolve phosphatases with Michaelis constants (K_M's) so large that they essentially cannot be saturated by their substrates. We showed that, in this scenario, the substrates can no longer respond ultrasensitively to incoming signals, and the phosphatases become highly inefficient enzymes that must be ex-pressed at high levels to ensure rapid substrate responses. A second mechanism that cells might employ involves disposing of specific phosphatases altogether, and instead employing degradation of the substrate as a means of removing phosphorylated molecules from the system. This alternative approach to "effective dephosphorylation" would have the benefit of reducing the required number of targets per phosphatase, decreasing the potential for phosphatase coupling in the rest of the network. We found that this dephosphorylation mechanism also cannot generate ultrasensitive responses. Additionally, in order for degradation to yield rapid response kinetics, the phosphorylated substrate would have to be highly unstable, with half-lives on the order of tens of minutes, which would involve high energetic costs to the cell.

Phosphatase promiscuity is likely a larger problem for serine/threonine phosphatases than it is for tyrosine phosphatases [13, 24, 78–80]. Interestingly, serine/threonine phosphatases like PP2A often act as holoenzymes comprised of a catalytically active subunit, a scaffolding subunit, and an "adaptor" subunit that recruits specific substrates to the complex (3, 28-34). Using our models, we demonstrated that these adaptor subunits can insulate signaling pathways from phosphatase crosstalk while still allowing each independent substrate to respond ultrasensitively. We found that the ability of adaptor subunits to insulate signals between different substrates depends upon the manner of adaptor binding: in particular, the adaptor must be able to bind the substrate independently of whether or not it is already bound in an active holoenzyme complex with the catalytic subunit. Focusing on the example of PP2A, it is likely that the substrate specificities of its adap-

tor subunits have evolved in order to functionally couple subsets of targets within the signaling network [13, 80–85].

Overall, our work demonstrates that, while there are mechanisms that can allow cells to avoid widespread phosphatase crosstalk, those mechanisms each involve a set of functional trade-offs that likely dictate which mechanism has evolved in any given situation. While these mechanisms almost certainly reduce the overall level of crosstalk in the cell, our analysis of the PP2A example indicates that at least some phosphatase coupling remains. Characterizing the functional role of phosphatase-mediated crosstalk in shaping network dynamics represents a major experimental and theoretical challenge in systems biology.

3.2 Materials and Methods

Our models of 2-Kinase/1-Phosphatase dynamics, the corresponding systems of Ordinary Differential Equations (ODEs), and details on the simulations are described in the Supporting Information. We used the CVODE library from SUNDIALS [86] to numerically integrate the systems of ODEs. Analytical solutions and subsequent derivations may also be found in the Supporting Information.

The half-lives of signaling proteins were taken from a published dataset of the half-lives of proteins in mouse C2C12 cells [87]. We analyzed the UniProt entry for each of the proteins in this dataset and checked for GO annotations describing the protein as being a phosphoprotein. The density curve for signaling protein half-lives was also obtained using the default density estimator in R [88].

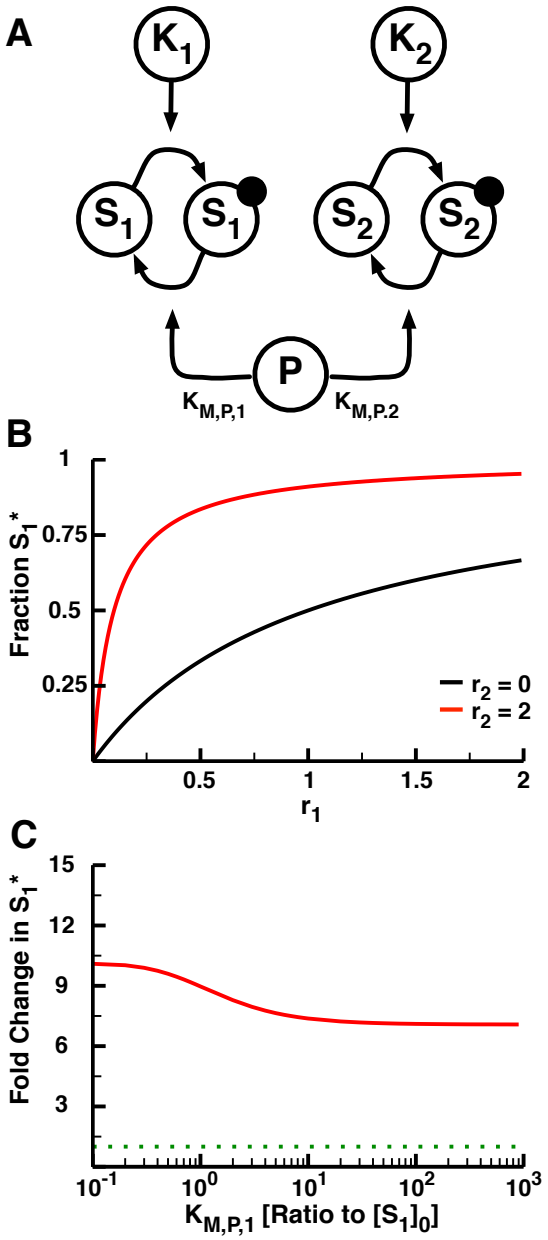

Figure 3.1: The 2-Kinase/1-Phosphatase Loop. (*A*) Two independent kinases (K_1 and K_2) phosphorylate their respective substrates (S_1 and S_2). The filled circles indicate the phosphorylation of the substrates. A single shared phosphatase P dephosphorylates both substrates. (*B*) The fraction of phosphorylated S_1 (denoted as S_1^*) as a function of response parameter r_1 when $r_2 = 0$ (black) and $r_2 = 2$ (red). r_1 and r_2 represent the ratio of the maximum velocity of the respective kinase to the maximum velocity of the phosphatase, and are the dominant response parameters for the system. The initial concentrations of both S_1 and S_2 are set at 10 μM, $K_{M,P,1}$ at 1 mM, and $K_{M,P,2}$ at 1 μM. As such, S_1 does not saturate either the kinase K_1 or phosphatase, whereas S_2 saturates K_2 and P. Note the increase in phosphorylated S_1 in response to activation of the second loop, as described previously for this motif [72]. (*C*) The fold change in S_1^* as a function of $K_{M,P,1}$. The fold change in S_1^* is calculated as the fraction S_1^* at $r_2 = 2$ divided by the fraction S_1^* at $r_2 = 0$. The dotted green line represents no change in S_1^*.

3.3 Results

3.3.1 The promiscuity of phosphatases

As mentioned above, kinases vastly outnumber phosphatases in the human genome [64, 65]. In order to characterize the generality of this kinase/phosphatase mismatch across different species, we searched the UniProt database for ratios of kinases to phosphatases and phosphoproteins to phosphatases (See Appendix B for details) [89]. We found that for most eukaryotes, there is no way to achieve a single, independent phosphatase per kinase, let along substrate (as is often assumed, if implicitly, in modeling studies) [80]. These findings are consistent with a variety of experimental studies, in which phosphatases have been shown to target tens to hundreds of phosphoproteins [13, 24, 78–80].

We previously demonstrated that phosphatases acting on multiple substrates could contribute to network crosstalk. Using a mathematical model in which a phosphatase is shared between two substrates with independent kinases (diagrammed in Fig. 3.1*A*), one can show that the shared phosphatase couples the responses of the substrates so that activation of one kinase increases the phosphorylation of both substrates through phosphatase inhibition [72]. It is straightforward to derive the fraction of phosphorylated substrate S_1 at steady state for this system:

$$S_1^* = \frac{(r_1-1)-(K_{K,1}+r_1\alpha_{P,1}K_{P,1})+\sqrt{((r_1-1)-(K_{K,1}+r_1\alpha_{P,1}K_{P,1}))^2+4(r_1-1)r_1\alpha_{P,1}K_{P,1}}}{2(r_1-1)} \tag{3.1}$$

where $S_1^* = [S_1^*]/[S_1]_0$ is the mole fraction of phosphorylated substrate S_1, $K_{K,1} = K_{M,K,1}/[S_1]_0$ and $K_{P,1} = K_{M,P,1}/[S_1]_0$ are the Michaelis constants of the substrate S_1 for the kinase K_1 and the shared phosphatase P divided by the total concentration of S_1, and $r_1 = k_{cat,K,1}[K_1]_0/k_{cat,P,1}[P]_0$. Since protein concentrations remain constant over relevant timescales, r_1 serves as the response parameter of that drives S_1 phosphorylation [41, 72]. The $\alpha_{P,1} = 1 + [S_2^*]/K_{M,P,2}$ term represents the influence of S_2 on the phosphorylation of S_1. The solution for the fraction of phosphorylated

S_2 at steady state is the same as Eq. 3.1 with different indices (e.g. $r_2 = k_{cat,K,2}[K_2]_0/k_{cat,P,2}[P]_0$). Upon activation of the second loop ($K_2 - S_2$), the phosphorylated S_2 acts as a competitive inhibitor of the phosphatase, increasing $\alpha_{P,1}$ if $[S_2^*]$ is large relative to $K_{M,P,2}$. This phosphatase inhibition results in an increase in S_1 phosphorylation (Fig. 3.1*B*). The difference in S_1 phosphorylation due to $K_2 - S_2$ activity can be illustrated as the fold change in S_1 phosphorylation upon K_2 stimulation, defined here as the concentration of phosphorylated S_1 at $r_2 = 2$ divided by the concentration of phosphorylated S_1 at $r_2 = 0$. Give this definition, we observed up to a 10-fold increase in phosphorylation of the first substrate at low values of r_1. Additionally, making P a poor phosphatase for S_1 by increasing $K_{M,P,1}$ does not remove the crosstalk, since $\alpha_{P,1}$ depends only on the saturation of the phosphatase by S_2^* (Fig. 3.1*C*). Although the results in Fig. 3.1 focus on a case where there is a single competing substrate S_2, multiple substrates can collectively saturate the phosphatase, leading to indirect activation of S_1 even when none of the competing substrates is at high enough concentration to individually saturate the enzymes (See Appendix B for details) [72].

3.3.2 Removing coupling with unsaturatable phosphatases

Since phosphatase coupling is dependent upon the collective saturation of the phosphatase by its substrates, it follows that an unsaturatable phosphatase could insulate substrate responses. To investigate the effects of phosphatase saturation on crosstalk we simultaneously increased both $K_{M,P,1}$ and $K_{M,P,2}$ in our 2-Kinase/1-Phosphatase model (Fig. 3.1*A*). When the $K_{M,P}$'s are smaller than the total concentrations of the substrates we see that the phosphorylation of S_1 at a low value of r_1 is increased about 10-fold upon activation of the second kinase. As the phosphatase K_M's are increased, however, the fold increase in S_1 phosphorylation drops until it reaches 1, indicating that S_1 becomes insensitive to K_2 activity at $K_{M,P}$'s above ~10 times total substrate concentration (Fig. 3.2*A*).

The insulation provided by an unsaturated phosphatase comes at the cost of the loss of an ultrasensitive response of the substrates to incoming signal [37, 72]. The unsaturated phosphatase can no longer operate at its maximum velocity, and as such it takes very little active kinase to phos-

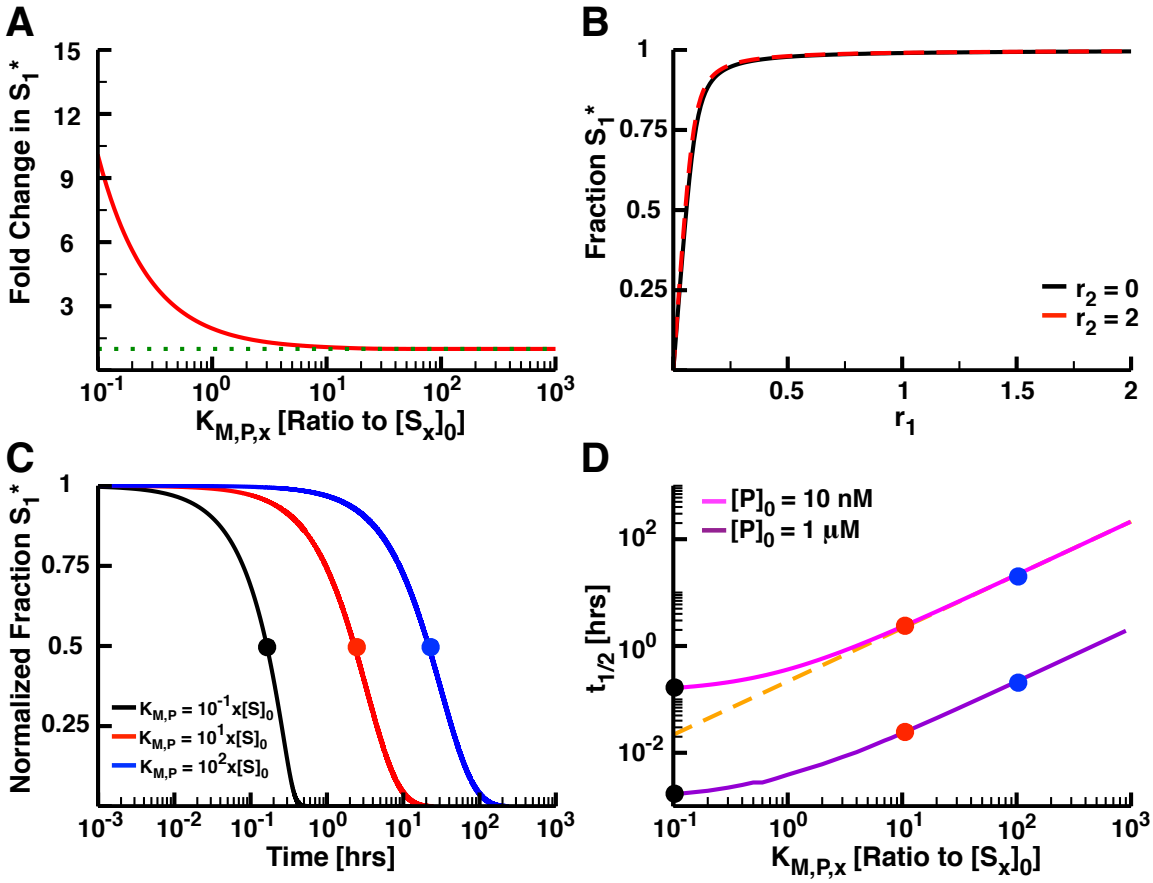

Figure 3.2: Removing coupling with unsaturatable phosphatases. (*A*) The fold increase in S_1 phosphorylation as a function of the $K_{M,P}$ of the shared phosphatase for both substrates in a 2-Kinase/1-Phosphatase loop. (*B*) The fraction S_1^* as a function of r_1 at $r_2 = 0$ (black) and $r_2 = 2$ (red) when $K_{M,P} = 10 \times [S]_0$. Note that there is very little difference between these curves. (*C*) The normalized fraction of phosphorylation substrate S_1 as a function of time after the removal of input signal. In these simulations, the concentration of S_2 and K_2 are set to 0. The systems were allowed to run to steady state at high K_1 activity ($r_1 = 2$); at $t = 0$, the activity of the kinase was set to 0 (i.e., $r_1 = 0$). The y-axes were normalized by y_1 = (y - min y) / (max y - min y), where min y is the fraction S_1^* at $r_1 = 0$ and max y is the fraction S_1^* at $r_1 = 2$ at steady state. (*D*) The half-life of S_1 phosphorylation as a function of $K_{M,P}$ with two total concentrations of P (10 nM, green, and 1 μM, purple). Note that the black, red, and blue dots are shown to illustrate the relationship between panels *D* and *C*. The dashed orange line shows the linear approximation of $t_{1/2}$ for highly unsaturated phosphatases ($t_{1/2} = K_{M,P}/k_{cat,P}[P]_0$).

phorylate a significant fraction of S_1 at steady state (Fig. 3.2*B*). These results can be understood by treating the system analytically. With an unsaturated phosphatase, the solution for the fraction of phosphorylated substrate becomes:

$$S_1^* = \frac{(1 + K_{K,1} + r_1 K_{P,1}) - \sqrt{(1 + K_{K,1} + r_1 K_{P,1})^2 - 4 r_1 K_{P,1}}}{2} \tag{3.2}$$

Note the lack of the α inhibition term from the shared phosphatase; since $K_{M,P,2} \gg [S_2]_0$, $\alpha_{P,1} \approx 1$, resulting in the lack of phosphatase-mediated crosstalk. One can show that the analytical solution in Eq. 3.2 for S_1^* is strictly hyperbolic in r_1 regardless of the values of the kinetic parameters, confirming the lack of any possible ultrasensitive response (see Appendix B for derivation) [37].

Another complication with an unsaturated phosphatase is the timescale on which it could dephosphorylate a pool of substrate molecules. To explore this issue, we initialized our system with a fully phosphorylated pool of S_1 molecules and no activity for the second kinase (i.e. $r_2 = 0$). This represents a pathway that has been fully activated by an incoming signal. We then ran this system with absolutely no kinase activity ($r_1 = 0$) to simulate the system after the removal of input. When $K_{M,P}$ is small it takes less than an hour to fully dephosphorylate all substrate. However, it takes longer as the phosphatase becomes unsaturated, taking over 100 hours to completely dephosphorylate the substrate when the $K_{M,P}$ is 100 times the total substrate concentration (Fig. 3.2*C*). When $K_{M,P}$ is large, the time it takes to dephosphorylate half the substrates ($t_{1/2}$) follows:

$$t_{1/2} = \frac{\log 2 \cdot K_{M,P,1}}{k_{cat,P}[P]_0} \tag{3.3}$$

The dependence of $t_{1/2}$ on $K_{M,P,1}$ is thus linear, which can result in very long timescales when the enzyme becomes highly unsaturated. In cases where fast dephosphorylation of the substrate is important (say, in tightly controlling the duration of a cellular response, or when fast oscillations are necessary), the system can compensate for this increase in timescale by expressing more phosphatase (Fig. 3.2*D*).

Interestingly, overexpression of the phosphatase alone can insulate substrate responses even

when the K_M's are small, so long as the concentration of phosphatase becomes high enough that the traditional Michaelis-Menten assumption that enzymes are at much lower concentration than their substrates is broken [16]. In this regime, however, the enzymes tend to sequester their substrates, reducing the concentration of unbound, phosphorylated substrate available to participate in downstream reactions within the network (see Fig. B.1 in Appendix B). Thus, while phosphatase overexpression can result in insulation and reduce dephosphorylation timescales (Fig. 3.2*D*), it can also reduce the capacity of the system to respond to incoming signals. As a result, many phosphatases (such as Msg5, which dephosphorylates the MAPK Fus3 in yeast) are at least an order of magnitude lower in concentration than their substrates [90, 91].

An increased phosphatase K_M is the only mechanism cells could use to desaturate the phosphatase. Previous experimental studies have demonstrated that binding of phosphorylated sites by SH-2 domain-containing or 14-3-3 proteins can shield the phosphoprotein from dephosphorylation, creating a reservoir phosphorylated substrate [92, 93]. The presence of such proteins would effectively reduce the concentration of substrate available to the phosphatase (and thus its saturation level) without influencing the saturation of the kinase. To consider the effects of this shielding phenomenon, we added substrate-specific "reservoir" proteins to our 2-Kinase/1-Phosphatase model. As expected, these reservoir proteins do insulate the substrates from crosstalk even when the K_M's are relatively small (see Figs. B.2*A* and *B* in Appendix B). However, since the phosphatase is effectively unsaturated, the substrates always respond hyperbolically to inputs. Dephosphorylation kinetics are also very slow in the presence of high concentrations of reservoir proteins, since the phosphatase must essentially "wait" for a substrate to be unbound long enough so that it can bind and catalyze the dephosphorylation reaction (Fig. B.2*C*).

Thus, while unsaturating the phosphatase K_M can insulate the response of substrates from one another, this mechanism clearly involves a set of trade-offs. For one, none of the substrates can respond in a switch-like manner to incoming signals, and so this mechanism cannot be deployed in cases where ultrasensitive responses are crucial [25, 42–46]. In addition, achieving fast dephosphorylation timescales may require high levels of phosphatase expression, which may become

impractical (or limit the capacity of the system to respond at all) in cases where the K_M needed to achieve insulation is very large (Figs. 3.2*D* and B.1*A*). These trade-offs likely limit the set of cases where insulation via an unsaturated phosphatase represents an evolutionarily favored mechanism.

3.3.3 Degradation as a phosphatase substitute

The vast majority of work on modeling signal transduction has assumed that dephosphorylation occurs through the catalytic activity of a phosphatase [25, 26, 44, 67, 69, 72–75]. While there are many clear cases where phosphatases play this role, it is also possible for substrate degradation to serve as a dephosphorylation mechanism. The idea in this case is straightforward: phosphoproteins are synthesized in their unphosphorylated state, but both the unphosphorylated and phosphorylated states of the protein may be lost from the system due to degradation. When the total protein concentration (regardless of state) remains constant in time, the effect of synthesis and degradation effectively amounts to a first-order dephosphorylation term (see Appendix B for details). Substrates that rely on degradation as their "phosphatase:" would not need a separate phosphatase enzyme, thus reducing the number of substrates each phosphatase would have to act upon.

In order to characterize degradation as a phosphatase substitute, we built a mathematical model with a substrate responding to a single kinase without a phosphatase. We assume the degradation process to be completely unsaturatable. As a result, the degradation terms are all taken to be first order, and there is no degradation- or phosphatase-mediated crosstalk between pathways. In some cases, phosphorylation of a protein changes its half-life; for instance, the phosphorylated state of the protein may be less stable than the unphosphorylated state [94, 95]. To capture this possibility in our model, the unphosphorylated and phosphorylated substrates are degraded at different rates ($k_{deg,U}$ and $k_{deg,P}$ for the unphosphorylated and phosphorylate states, respectively). When the substrate is bound to a kinase, we assume that the kinase falls off the complex as the substrate is degraded. Unphosphorylated substrate is synthesized at a rate necessary to maintain a constant total substrate concentration at steady state (see Appendix B for details). In order to parameterize the model, we obtained a range of half-lives by using UniProt to identify phosphoproteins from

published dataset of protein half-lives in mouse C2C12 cells [87, 89]. Phosphoproteins in this dataset have half-lives ranging from about 10 to 187 hours, with a median of 31 hours (Fig. 3.3*A*). We used these values to set the range of biologically relevant degradation rates for the substrates.

This model was then run to steady state using each of the phosphoprotein half-lives for a range of kinase concentrations. For these initial simulations we assumed that the phosphorylation state does not influence degradation rate (i.e. $k_{deg,U} = k_{deg,P}$). The response of the substrate to incoming signal is dependent upon the half-life of the substrate; more stable phosphoproteins are more highly phosphorylated than less stable phosphoproteins (Fig. 3.3*B*). However, substrates with any half-life in the mouse data set become completely phosphorylated when the kinase concentration is at or above 1 nM, making these substrates highly sensitive to incoming signal. Note that catalytic rate of the kinase in these simulations is 0.9 s^{-1}, which is close to experimentally determined kinase catalytic rates that have been determined experimentally [96, 97]. Faster kinase catalytic rates would further reduce kinase concentrations necessary to completely phosphorylate substrates with the observed half-lives.

To understand how the different degradation rates may affect substrate phosphorylation and sensitivity, we derived a solution for the fraction of phosphorylated S^* in this model:

$$S^* = \frac{1 + r_{deg} + K_{deg} - \sqrt{(1 + r_{deg} + K_{deg})^2 - 4r_{deg}}}{2} \tag{3.4}$$

where $r_{deg} = k_{cat,K}[K]_0/k_{deg,P}[S]_0$ is the ratio of the maximum velocity of the kinase to the maximum velocity of substrate degradation. This solution includes a modified Michaelis constant $K_{M,deg} = (k_{-,K} + k_{cat,K} + k_{deg,U})/k_{+,K}$, taking into account the degradation of the kinase-bound substrate, which is divided by the total substrate concentration to obtain K_{deg}. As expected based on our results for an unsaturatable phosphatase, one can show that the phosphorylation response of the substrate in this model can only be hyperbolic in r_{deg} (See Appendix B for derivation). As such, a signaling pathway that relies on degradation to remove phosphorylated substrate could never respond ultrasensitively to incoming signals.

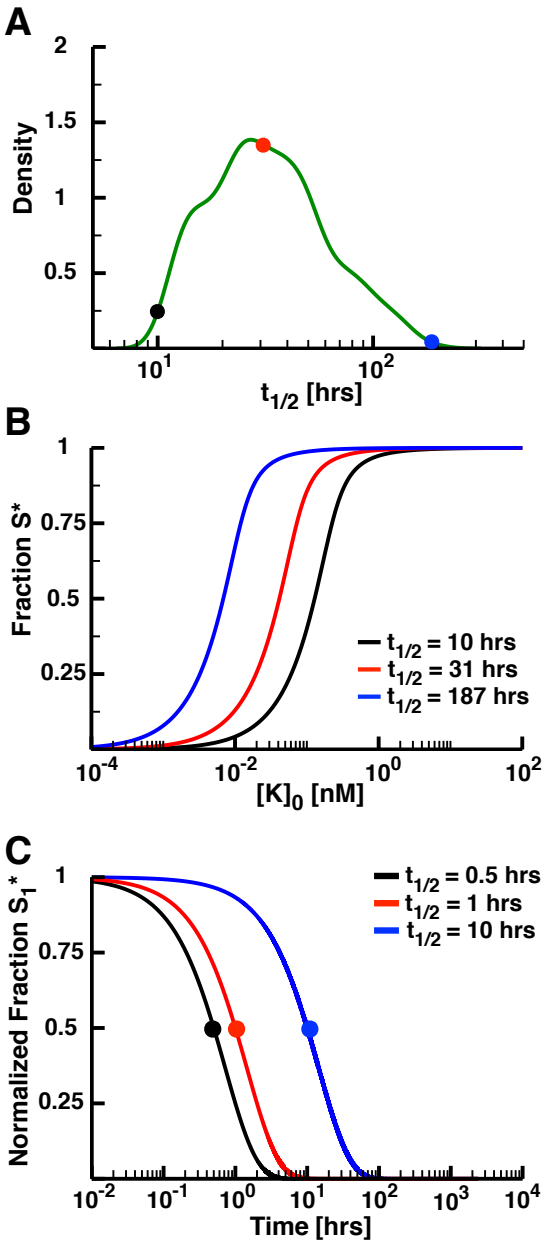

Figure 3.3: Degradation as a phosphatase. (*A*) The range of signaling protein half-lives in mouse C2C12 cells [87]. The minimum reported half-life is 10 hours (black dot), maximum is 187 hours (blue), and median is 31 hours (red). (*B*) The fraction S^* as a function of the concentration of active kinase, $[K]_0$. Note that the model for these panels includes only the kinase and degradable substrate. $[K]_0$ directly affects the parameter r_{deg}, which represents the ratio of the maximum velocity of the kinase to the maximum velocity of degradation ($k_{deg} \times [S]_0$). The value of r_{deg} will thus change for a constant concentration of kinase as the half-life changes. As such, any r_{deg} for a protein with a longer half-life involves a lower concentration of kinase than for a shorter half-life. (*C*) The normalized fraction S^* as a function of the time after setting kinase activity to 0 for substrates with half-lives of 0.5, 1, and 10 hours.

Finally, the timescales required to degrade a pool of phosphorylated substrate based upon the half-lives of phosphoproteins from the C2C12 dataset would be very long. To demonstrate this, we ran the model at high levels of kinase activity ($r_{deg} = 2$) to steady state. We then set $r_{deg} = 0$ to simulate the system after the removal of the input. Even with the shortest substrate half-life, the system requires about 100 hours to completely dephosphorylate the substrate (Fig. 3.4*C*). Complete dephosphorylation of the substrate in less than 1 hour would require that the phosphorylated state have a half-life on the order of minutes. If the system needs to recover quickly from incoming signals, then utilizing degradation as a phosphatase subunit would likely be quite inefficient, requiring a very unstable phosphorylated substrate. Maintaining a reasonable concentration of total substrate would, in turn, require a high rate of protein synthesis, resulting in a high energetic cost for the cell. Thus, while degradation could reduce the total "substrate burden" of phosphatases in the cell, it is likely to be employed only in cases where an ultrasensitive response is not necessary and when either slow dephosphorylation kinetics, or the energetic costs of high protein turnover, are functionally acceptable.

3.3.4 Role of phosphatase regulatory subunits in pathway isolation

Serine phosphatases like PP1 and PP2A exist as holoenzymes: the catalytic subunit of PP2A binds to a scaffold subunit, which makes up the catalytic core. This core can then bind to one of many possible regulatory adaptor subunits, with each adaptor recruiting PP2A to a specific set of substrates (Fig. 3.4*A* [24, 76, 80–84]. The mechanism of targeting of substrates by a regulatory subunit and the effects of concentration and binding constants has been examined previously [98]. However, this study did not consider whether regulatory subunits might provide insulation from phosphatase-mediated crosstalk. In order to investigate the role of these adaptors in isolating different pathways that share the same phosphatase catalytic core, we built two models, an "ordered" and an "unordered" model for the binding of the phosphatase adaptor subunits. In the ordered model, the phosphatase catalytic core can bind two one of two adaptor subunits, creating two distinct holoenzymes. This holoenzyme can then specifically bind to and dephosphorylate a cognate

substrate (diagrammed in the inset of Fig. 3.4*B*). This model represents the standard picture where the holoenzyme first assembles and then acts on its substrates [82].

We first tested the ordered model to determine if the interaction specificity of adaptor subunits is sufficient to isolate the responses of different substrates. In order to examine how concentrations of the adaptor subunits affect substrate phosphorylation, we ran this model to steady state with a range of total adaptor concentrations (keeping the concentrations of the two adaptors equivalent, i.e. $[A_1] = [A_2]$) and measured the fraction of phosphorylated S_1 at low levels of kinase 1 activation ($r_1 = 0.05$) and high levels of kinase 2 activity ($r_2 = 2$) (Fig. 3.3*B*). When the concentration of the adaptors is less than the concentration of phosphatase catalytic core, the phosphorylation of S_1 is very high. In this model, a phosphatase can only act on its substrate when bound to an adaptor subunit. When the concentration of adaptors is less than that of the catalytic core, the total number of active phosphatases is thus limited by the adaptor concentration. As that concentration decreases, so does the concentration of active phosphatases, making the apparent value of r very high for both substrates, and leading to an increase in S_1 phosphorylation (Fig. 3.4*B*). Once the concentration of the adaptors exceeds that of the catalytic core, however, the concentration of the adaptors has little influence on S_1 phosphorylation (Fig. 3.4*B*). At higher levels of kinase 1 activity, however, adaptor subunits provide no insulation between different substrates in this model; S_1 phosphorylation increases considerably as r_2 goes from 0 to 2 (Fig. 3.4*C*). This indicates that, while regulatory adaptor domains may help to target different phosphoproteins, their specificity cannot insulate substrates from phosphatase-mediated crosstalk in the ordered model. This is because the large pool of phosphorylated S_2 makes it far more likely that the holoenzyme will remain intact in order to bind and further dephosphorylate S_2. As such, the catalytic core is prevented from disassociating its S_2-specific adaptor subunit and forming the holoenzyme specific to S_1, decreasing the concentration of phosphatase available to act on S_1.

The unordered model differs in that the adaptor subunit can bind its specific phosphoprotein without being bound to the catalytic core first. The adaptor-substrate dimer can then bind the catalytic core, resulting in substrate dephosphorylation (diagrammed in the inset of Fig. 3.5*A*). This

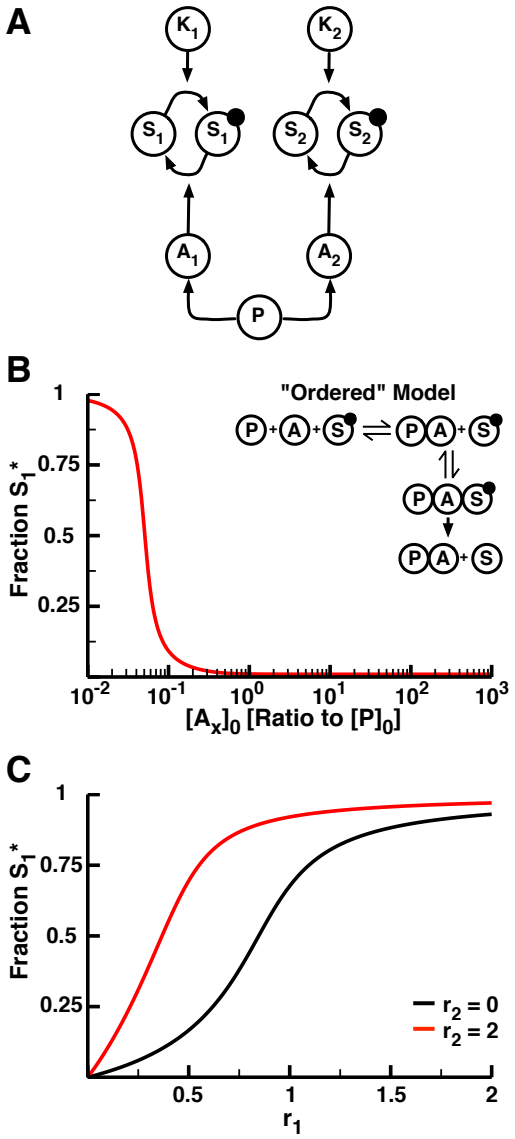

Figure 3.4: The "ordered" model of phosphatase regulatory adaptor subunits. (*A*) Diagram of the regulatory adaptor model. The previously described 2-Kinase/1-Phosphatase model (Fig. 3.1*A*) was expanded to include two regulatory adaptor subunits, A_1 and A_2, which target the phosphatase to substrates S_1 and S_2, respectively. The phosphatase must bind an adaptor domain in order to bind and dephosphorylate either substrate. (*B*) The fraction S_1^* as a function of the initial concentration of each adaptor subunit with low levels of K_1 activity (r_1 = 0.05) and high levels of K_2 activity (r_2 = 2). The concentrations of both adaptor subunits are equal and are varied simultaneously. *Inset*: Diagram of the ordered model. The phosphatase must first bind to an adaptor subunit. This complex then can then bind to and dephosphorylate a phosphorylated substrate. (*C*) The fraction S_1^* as a function of r_1 at $r_2 = 0$ (black) and $r_2 = 2$ (red) with $[A_x]_0 = 100$ nM. In this model the adaptor subunits do not eliminate the potential for phosphatase-mediated crosstalk, despite being perfectly specific for the respective substrates.

removes the ability of one substrate to sequester the catalytic core in the holoenzyme since the core can dissociate from the regulatory adaptor subunit even in the presence of the phosphorylated substrate. At low concentrations of the adaptor subunits, the unordered model acts similar to the ordered model (Fig. 3.5*A*). However, as the adaptor concentration becomes very large there is an increase in substrate phosphorylation. This is due to the prozone effect; as the concentration of adaptor increases the system starts to produce many phosphatase-adaptor dimers and phosphorylated substrate-adaptor dimers. These dimers cannot bind to one another, thus preventing the catalytic core from interacting with the substrate [91, 98–103]. Even though the prozone effect can influence the response at high adaptor concentration, there is still a wide range of adaptor concentrations (over two orders of magnitude) that provide robust phosphatase activity (Fig. 3.5*A*). The unordered model also provides effective pathway insulation: there is essentially no change in S_1 phosphorylation as the second kinase switches between inactive and active ($r_2 = 0$ and $r_2 = 2$, respectively) (Fig. 3.5*B*). Additionally, S_1 can respond ultrasensitively in r_1, although the apparent r_1 is about half of what is expected since the adaptor subunits split the available phosphatase between the two substrates.

PP2A is known to interact with a large number of substrates, and yet only about 18 different regulatory subunits have been identified. In order to understand how this handful of subunits could allow the PP2A catalytic core to dephosphorylate so many substrates while maintaining a degree of pathway insulation, we built a network based upon previously identified interactions between the regulatory subunits and the substrates of PP2A (Fig. 3.5*C*) [76]. This network depicts the interactions between 42 PP2A substrates and the 18 regulatory subunits. Strikingly, only 6 of the substrates have interactions with more than one of the adaptors, and of these, only p53 interacts with three. The remaining substrates are all specific to a single regulatory subunit. According to our findings, these subsets of phosphoproteins would be insulated from one another, assuming the regulatory subunits can bind the phosphorylated substrates without binding the catalytic core. However, this insulation would not work within any given subset: since they all share a single regulatory adaptor, that subset would act as if they have a single, shared phosphatase. The coupling

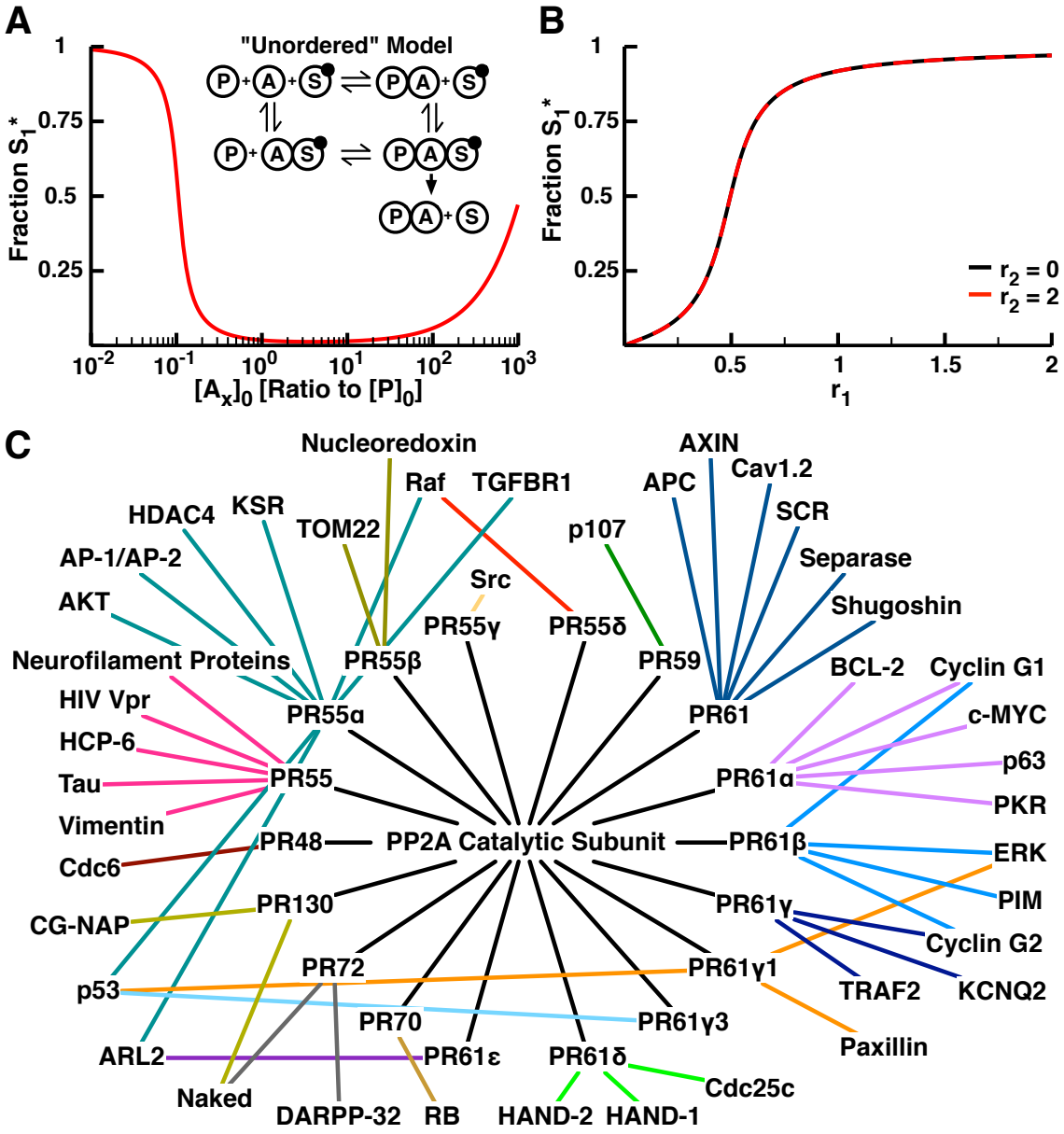

Figure 3.5: The "unordered" model of phosphatase regulatory adaptor subunits. (*A*) The fraction S_1^* as a function of initial concentration of each adaptor subunit with low levels of K_1 activity (r_1 = 0.05) and high levels of K_2 activity ($r_2 = 2$). *Inset*: Diagram of the unordered adaptor model. In this case, the adaptor subunit can bind to the substrate without previously binding to the phosphatase (lower path). (*B*) The fraction S_1^* as a function of r_1 at $r_2 = 0$ (black) and $r_2 = 2$ (red) with $[A_x]_0 = 100$ nM. Note that this model provides insulation between the substrates. (*C*) The network diagram of the interactions of the PP2A catalytic subunit with each of 18 identified regulatory subunits [76]. The PP2A holoenzyme formed by binding a regulatory subunit then interacts with a subset of the proteins in the network.

within groups of substrates could have positive phenotypic effects. For instance, PR55α interacts with AKT, AP-1/AP-2, HDAC4 and KSR, each of which are involved in the same signaling cascade [104–107]. We have previously shown that sharing a phosphatase among multiple members of the same signaling cascade increases the sensitivity of the cascade [72], so sharing the same PP2A holoenzyme within the a subsection of a signaling network could increase its local responsiveness to external signals.

3.4 Discussion

While it is understood that both kinases and phosphatases act on large number of targets, phosphatases have classically been considered more "promiscuous". For example, while there have been more than 20 phosphoproteins identified as substrates of the EGF/ErbB receptor tyrosine kinase [108], phosphatases like PP2A dephosphorylate hundreds of substrates within the cell [76–78]. Since phosphatases must act on large numbers of substrates, it is unclear how cells can insure at least some degree of specificity in downstream responses to incoming signals. In this work, we extended our previous modeling efforts to consider three possible mechanisms whereby cells might insulate substrates from the possibility of phosphatase crosstalk. We found that each mechanism involves a set of unavoidable functional trade-offs that likely influence where and when they have evolved in eukaryotic signaling networks.

Perhaps the simplest approach would be to evolve phosphatases whose K_M's are so large that they cannot be saturated by their substrates. While this approach is indeed effective in isolating the responses of substrates that share a particular phosphatase (Figs. 3.2*A* and *B*), this mechanism inherently creates phosphatases that are highly inefficient enzymes. One result of this fact is that the response of a substrate to upstream signals can no longer exhibit a switch-like, ultrasensitive character. It is thus unlikely that phosphatase inefficiency would be employed in cases where ultrasensitivity is a key component of the functional response [16, 42, 43, 45, 46, 64, 69, 72]. A second consequence of this inefficiency is the fact that dephosphorylation kinetics can be slow,

which would require high phosphatase expression levels in order to allow a particular substrate to quickly return to baseline activity levels after a signal is removed (Figs. 3.2*C* and *D*). Other mechanisms that effectively unsaturate the phosphatase, such as the expression of high levels of "reservoir" proteins that bind the phosphorylated state and prevent phosphatase binding, can also provide insulation, but with similar impacts on ultrasensitivity and dephosphorylation

Protein degradation can also assume the role of an "effective phosphatase," reducing the number of substrates that phosphatase enzymes might have to act upon in the cell. If we assume that degradation is both efficient and unsaturatable, then degradation can indeed prevent phosphatase crosstalk between substrates. As with the inefficient phosphatase mechanism, however, the lack of an ultrasensitive response and slow dephosphorylation kinetics might represent an issue for this particular mechanism in some cases. Even the least stable phosphoprotein in mouse cells would still require over 100 hours to return to baseline after removal of the signal (Fig. 3.3*C*). Using degradation to fill the role of a phosphatase would thus require a very high (and very costly) rate of protein turnover in cases where fast response kinetics are necessary.

The final mechanism we considered here involved having separate "regulatory subunits" that recruit a catalytic core to particular substrates. While these subunits can clearly provide substrate specificity, it is not clear that they can actually insulate those substrates from one another. Indeed, if the phosphatase pre-assembles into a holoenzyme before interacting with the substrate (which is essentially the classical picture for this sort of enzyme [24, 76, 80–84]), then even perfectly specific regulatory subunits cannot prevent phosphatase-mediated crosstalk (Fig. 3.4). If the regulatory subunits can bind their substrates independently of binding the catalytic core, however, then this mechanism can provide insulation while still maintaining the possibility of an ultrasensitive response, and fast dephosphorylation kinetics (Fig. 3.5). Deploying this mechanism, however, would require the evolution and expression of a distinct regulatory subunit for every set of phosphoproteins that the cell needs to isolate, which could represent a costly and evolutionarily complex solution to the problem of phosphatase coupling.

Our findings thus demonstrate that cells have a considerable degree of flexibility in the mecha-

nisms they might use to insulate substrates from one another despite the (relatively) small numbers of phosphatases in eukaryotic genomes. The question then becomes: which of these mechanisms are deployed in any given situation, and to what extent are substrates truly isolated from one another? For instance, the well-characterized serine/threonine phosphatase PP2A acts as a holoenzyme with a regulatory subunit [24, 76, 80–84], and it is currently unclear if the assembly of this holoenzyme follows the ordered or unordered model. Our results suggest that an experiment in which the adaptor concentration is increased (either directly in an in vitro setting or through overexpression in vivo) could establish which of these mechanisms is utilized by PP2A (Figs. 3.4*B* and 3.5*A*). Even if the assembly mechanism is unordered, PP2A-mediated crosstalk is still a strong possibility. While the data presented in Fig. 3.5*C* are certainly not complete (for instance, KSR and Akt share an additional regulatory subunit [109, 110]), PP2A clearly does not have a distinct regulator for every substrate with which it interacts. Instead, these regulators have clearly evolved to interact with a specific subset of proteins, possibly coupling their responses in functionally meaningful ways (Fig. 3.5*C*).

A major component of systems biology is the construction of formal mathematical or computational models of cellular regulatory systems, with the goal of understanding how cells process information from their environment and respond appropriately [111–114]. In the case of complex eukaryotic signaling networks, a major barrier to this goal is the fact that dephosphorylation, by phosphatases or through some other mechanism, has been comparatively poorly characterized for most phosphoproteins in the network. The addition of anonymous and perfectly specific phosphatases to cover this gap may produce effective models of individual cascades or pathways [25, 44, 67, 69, 73–75], but it is unlikely this practice will remain effective as larger, more genome-wide models of signaling networks are formulated. The experimental determination of the phosphatase structure of signaling systems, and the theoretical understanding how that structure has evolved to generate and regulate crosstalk among pathways (Fig. 3.5*C*), thus represents a major challenge for systems biology.

Chapter 4

Crosstalk and the Evolution of Specificity in Two-Component Signaling

4.1 Introduction

Two-component signaling (TCS) represents the primary signaling modality in bacteria [2]. The prototypical TCS pathway includes a membrane-bound sensor histidine kinase (HK) that autophosphorylates upon receiving an input signal. The HK then binds and transfers its phosphoryl group to a response regulator (RR), which often functions directly as a transcription factor, regulating gene expression patterns in response to the signal [2, 3]. Many HKs are bifunctional, acting as both the kinase and phosphatase for their RR; the ratio of kinase to phosphatase activity, and thus the phosphorylation state of the RR, is controlled by the input [2, 3, 27, 32, 33, 115–117].

Signaling networks in eukaryotes display extensive "crosstalk," with individual kinases acting on large numbers of targets: the kinase Cdk1, for instance, has hundreds of substrates in yeast [17, 60, 61]. Bacterial TCS networks show a remarkably different topology: HKs usually act on a single target [4–9]. Intensive experimental study over the past 10 years has revealed the biochemical and biophysical basis for this lack of promiscuity. In general, HKs demonstrate a strong "kinetic preference" for their cognate substrates, preferentially phosphorylating them on short timescales

[8, 9, 27–31]. A relatively small number of residues in the protein-protein interaction interface between HKs and RRs is responsible for maintaining this specificity [6, 8, 9, 30, 31, 118]. Recently, Capra *et al.* [119] demonstrated that making just two mutations in this interface could introduce an interaction between an HK (PhoR) and a noncognate RR (NtrX) in *Escherichia coli*. This exogenous interaction decreased phosphate starvation signaling, leading to profound decreases in growth rate and fitness in mutant cells grown under phosphate-limiting conditions. It has been shown that adding crosstalk to TCS can reduce information transfer efficiency under certain conditions [120], but it remains unclear exactly why TCS pathways are constrained from evolving crosstalk.

One of the most common motifs in eukaryotic signaling networks is a pair of enzymes (e.g., a kinase and a phosphatase) acting on a shared substrate (Fig. 4.1*A*) [26, 72]. Using mathematical models, we recently showed that adding multiple competing substrates to this type of Goldbeter-Koshland (GK) loop would tend to induce and ultrasensitive, switch-like behavior in the system, which could easily have positive phenotypic consequences for the cell [25, 42, 44, 72]. In the work described here, we performed a similar analysis, extending a well-studied and validated mathematical model of bifunctional HKs (Fig. 4.1*B*) to the case of multiple substrates [32, 33]. We found that, because the HK acts both as the kinase and the phosphatase in these systems, the addition of competing interactions with multiple RRs always decreases the response of the cognate RR. This is consistent with the findings of Capra *et al.* [119], who showed that the phenotypic effects of their crosstalk mutant were not due to the misregulation of NtrX targets, but rather a direct result of decrease in phosphate starvation signaling.

The pressure to maintain cognate signaling suggests the existence of a barrier in the evolution of new TCS pathways. New HK-RR pairs can arise from the duplication of existing HK-RR genes, which subsequently diverge into a new pathway [31, 121]. There is unavoidable crosstalk immediately postduplication, which can attenuate the response to the original signal. Using our models, we characterized a set of "near-neutral" evolutionary trajectories that minimize the impact of the new pair on the signaling of the parent pathway. All of these trajectories involved insulating the two pathways from one another before establishing new input and output functionalities. To

test this prediction, we, separately aligned multiple HK and input domain sequences from fully sequenced bacterial genomes. Analysis of the K_A/K_S ratios of the most recently diverged domains revealed that the interaction interface of the HK is under strong positive selection immediately after duplication, likely owing to the pressure to insulate interactions between the parent and duplicate pairs [8, 9, 27–31]. Input domains in the HKs often evolve through "domain-swapping," whereby a new pathway picks up input functionality b wholesale exchange of domains with other proteins in the genome [31, 121]. Analysis of K_S values indicates that these swapping events generally occur only after the HK interfaces have had sufficient time to evolve interaction specificity. These findings suggest that the majority of HK-RR duplications follow the near-neutral evolutionary paths we predicted. Overall, our work indicates that the bifunctional nature of HKs has likely been a major driving force in the evolution of insulated topologies in bacterial signaling networks [31].

4.2 Materials and Methods

Our model of TCS dynamics, and the corresponding systems of ordinary differential equations (ODEs), is described in Appendix C, section 1. We used the CVODE package from SUNDIALS [86] to numerically integrate the system of ODEs. Nucleic acid and amino acid sequences of HKs were obtained from the KEGG database [20], and domain boundaries were obtained from Pfam annotations [122]. The amino acid sequences of the domains within each genome were aligned using CLUSTALW [123]. The nucleic acid sequences were then mapped to the amino acid multiple sequence alignments. K_A and K_S values were obtained using the seqinR library in the R statistical computing platform [124]. Further details on our simulations and analyses can be found in Appendix C.

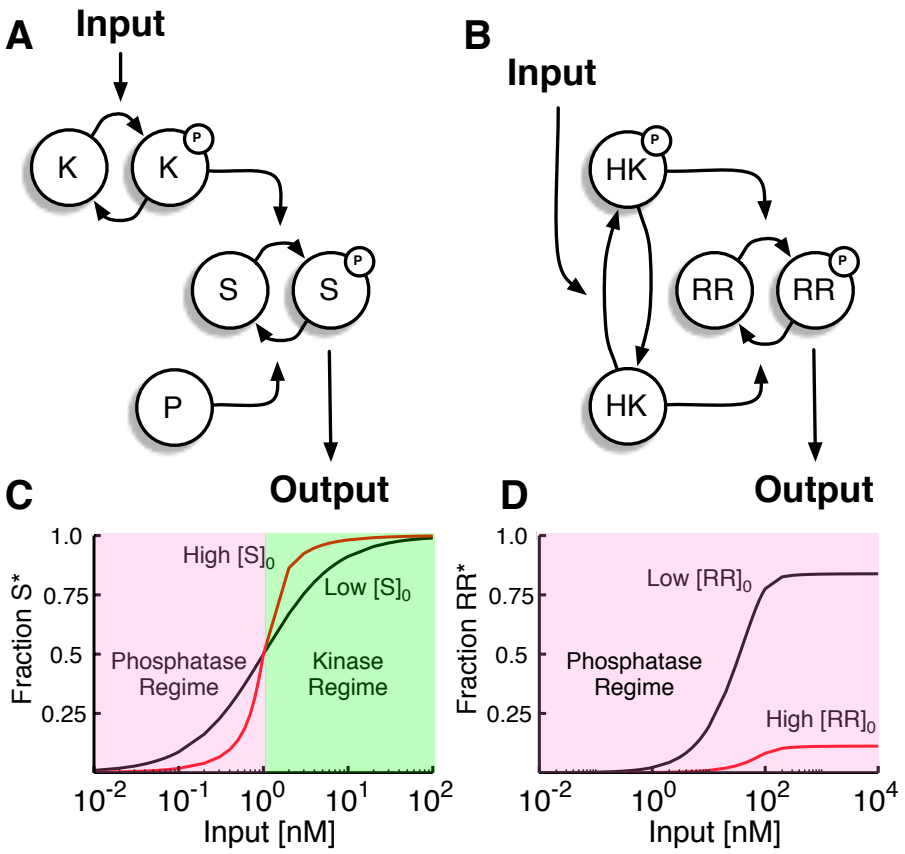

Figure 4.1: TCS pathways vs. Goldbeter-Koshland loops. (*A*) Diagram of a Goldbeter-Koshland loop. An input activates a kinase K, which phosphorylates a substrate. The phosphatase P is a separate enzyme that undoes this modification. (*B*) Diagram of a TCS pathway. An input causes the autophosphorylation of an HK, which transfers its phosphoryl group to the RR. The unphosphorylated HK also serves as the phosphatase. (*C*) The fraction of phosphorylated substrate S as a function of input concentration (on a log scale) for two total concentrations of S ($[S]_0 = 100$ nM, black and $[S]_0 = 10$ μM, red). The phosphatase regime and kinase regime defined in the main text are shaded pink and green, respectively. Note that the addition of substrate makes the response more switch-like [26]. (*D*) The fraction of phosphorylated response regulator RR as a function of input concentration for two total concentrations of RR ($[RR]_0 = 100$ nM, black, and $[RR]_0 = 10$ μM, red). As discussed in the text, HKs are always in the phosphatase regime, so the entire plot is shaded pink. Note that increasing total substrate concentration in this case reduces the response efficiency of the RR.

4.3 Results

4.3.1 Response to changes in RR concentration

To understand the impact of crosstalk on signaling, it is helpful to consider the response of the system to changes in the concentration of a single substrate [72]. As mentioned above, eukaryotic signaling networks are formed largely from a motif in which one enzyme (e.g., a kinase K) modifies a substrate and a second enzyme (e.g., a phosphatase P) removes the modification (Fig. 4.1A). Goldbeter and Koshland first characterized the behavior of this system over 30 years ago, finding that the response of the system at steady-state followed:

$$S^* = \frac{(r-1)-(K_K+rK_P)+\sqrt{((r-1)-(K_K+rK_P))^2+4(r-1)rK_P}}{2(r-1)} \tag{4.1}$$

where $S^* \equiv [S^*]/[S]_0$ is the mole fraction of phosphorylated substrate, $K_K \equiv K_{m,K}/[S]_0$ and $K_P \equiv K_{m,P}/[S]_0$ are the Michaelis constants divided by the total concentration of substrate, and $r \equiv k_{cat,K}[K]_0/k_{cat,P}[P]_0$ is the ratio of the maximum velocities of the enzymes [26]. Because protein concentrations (and thus the saturation parameters) remain constant over short timescales [41], r represents the dominant response parameter. In Fig. 4.1C, we considered a model of a GK loop in which an explicit input molecule binds and activates the kinase, thus modulating r (Appendix C). At unsaturating concentrations of substrate, substrate phosphorylation increases hyperbolically (Fig. 4.1C). At saturating concentrations, however, the system displays a switch-like behavior known as "0^{th}-order ultrasensitivity." When $r < 1$, phosphatase activity dominates and the addition of substrate decreases S^*; we call this the "phosphatase regime." When $r > 1$, kinase activity dominates and the addition of substrate increases S^*; this is the "kinase regime." These two opposing trends lead to an increasingly ultrasensitive response as total substrate concentration increases (Fig. 4.1C) [25, 26, 42, 44, 72].

A major difference between eukaryotic GK loops and bacterial TCS is the fact that the HK often acts as both kinase and phosphatase for its substrate RR (Fig. 4.1B). Ten years ago, Batchelor

and Goulian [32] developed and approximate analytical solution of a mathematical model of TCS signaling and demonstrated that the concentration of phosphorylated RR ($[RR^*]$) was insensitive to changes in total RR concentration ($[RR]_0$). To study crosstalk in TCS, we constructed a model very similar to that of Batchelor and Goulian and other authors (see Appendix C for the details of the model) [32, 33]. We were able to obtain a complete analytical solution in this case and found that the steady-state response of the system follows:

$$RR^* = \frac{(r\beta - \beta') - (\varepsilon' K_K + \varepsilon r K_P) + \sqrt{((r\beta - \beta') - (\varepsilon' K_K + \varepsilon r K_P))^2 + 4(r\beta - \beta')\varepsilon r K_P}}{2(r\beta - \beta')} \quad (4.2)$$

where $RR^* \equiv [RR^*]/[RR]_0$ is the fraction of phosphorylated response regulator, K_K and K_P are as previously defined, and $r \equiv k_{cat,K}/k_{cat,P}$ becomes a constant ratio between the catalytic rates of the kinase and phosphatase reactions. In this case, the dominant response parameter to changes in input are the new β and ε terms, which are dependent upon the autophosphorylation and autodephosphorylation rates of the HK and thus the input signal (see Appendix C for derivation and details). We compared the predictions of this solution to previous experimental results by Batchelor and Goulian [32] in which the concentrations of the HK and RR were varied and found that Eq. 4.2 reproduces their data (Appendix C).

As with the GK loop, we considered a case in which an explicit input molecule binds and activates the HK (Fig. 4.1*B*). Because bacterial TCS are well studied, experimental values are available for both total concentrations and kinetic parameters in this model (Appendix C) [3, 5, 32, 125]. Using those parameters for the purpose of display, we found a dramatic decrease in RR^* when total RR concentration is high (Fig. 4.1*D*). Note that this is the fraction of phosphorylated RR; the total concentration of active RR molecules does not depend on $[RR]_0$ when the RR is at saturating concentrations, as previously noted (Appendix C) [32]. Thus, although the response of the system is robust to changes in total RR in this regime, it also becomes inefficient; that is, increasing the expression of the RR does not increase the response capacity of the system.

In the GK loop (Fig. 4.1*A* and *C*), the separation of the kinase and phosphatase regimes depends

upon the term $(r-1)$ in the denominator of Eq. 4.1, which is negative in the phosphatase regime and positive in the kinase regime. Eq. 4.2 contains a similar term, $(r\beta-\beta')$, and this term is always negative. TCS loops are thus always in the phosphatase regime, and the general trend in Fig. 4.1*D* does not depend on specific values of kinetic parameters (Appendix C). This behavior ultimately arises from the fact that both the kinase and phosphatase reactions produce unphosphorylated HK, which itself is a phosphatase, keeping the system in the phosphatase regime.

4.3.2 Competition in TCS

To consider crosstalk in TCS, we added a single competing RR to the system diagrammed in Fig. 4.1*B*. We denote the cognate RR as RR_1, the noncognate interaction partner as RR_2, and define the ratio of their total concentrations to be $R \equiv [RR_2]_0/[RR_1]_0$. To account for the impact of crosstalk on RR_1 function, we also added explicit output molecules (O_1 binding phosphorylated RR_1 and O_1 binding phosphorylated RR_2) to the model. To study the responses of this system to a competing RR when the HK displays no kinetic preference for either substrate, we set the kinetic parameters of RR_2 to be the same as those for RR_1 [8, 9, 27–31]. We found that adding RR_2 at the same total concentration as the cognate substrate results in a decrease in output activity of the system, which we defined as the fraction of O_1 molecules bound by phosphorylated RR_1 (Fig. 4.2*A*). As the total concentration of RR_2 is increased, impact on RR_1 activity becomes even more significant. The situation is similar to that in Fig. 4.1*D*, but in this case the total concentration of RR_1 is constant, so both the fraction and concentration of phosphorylated RR_1 is constant, so both the fraction and concentration of phosphorylated RR_1 decreases, leading to a decrease in output activity. Our results thus indicate that the type of crosstalk introduced experimentally by Capra, *et al.* [119] into bacterial cells would likely decrease the performance of the PhoR/PhoB signaling system, providing an explanation for the lower fitness of crosstalk mutants in phosphate-limiting conditions.

Bacterial genomes can encode 5-200 HK-RR pairs, depending on the species in question [31]. To consider the impact of crosstalk in such cases, we expanded the model each HK can interact

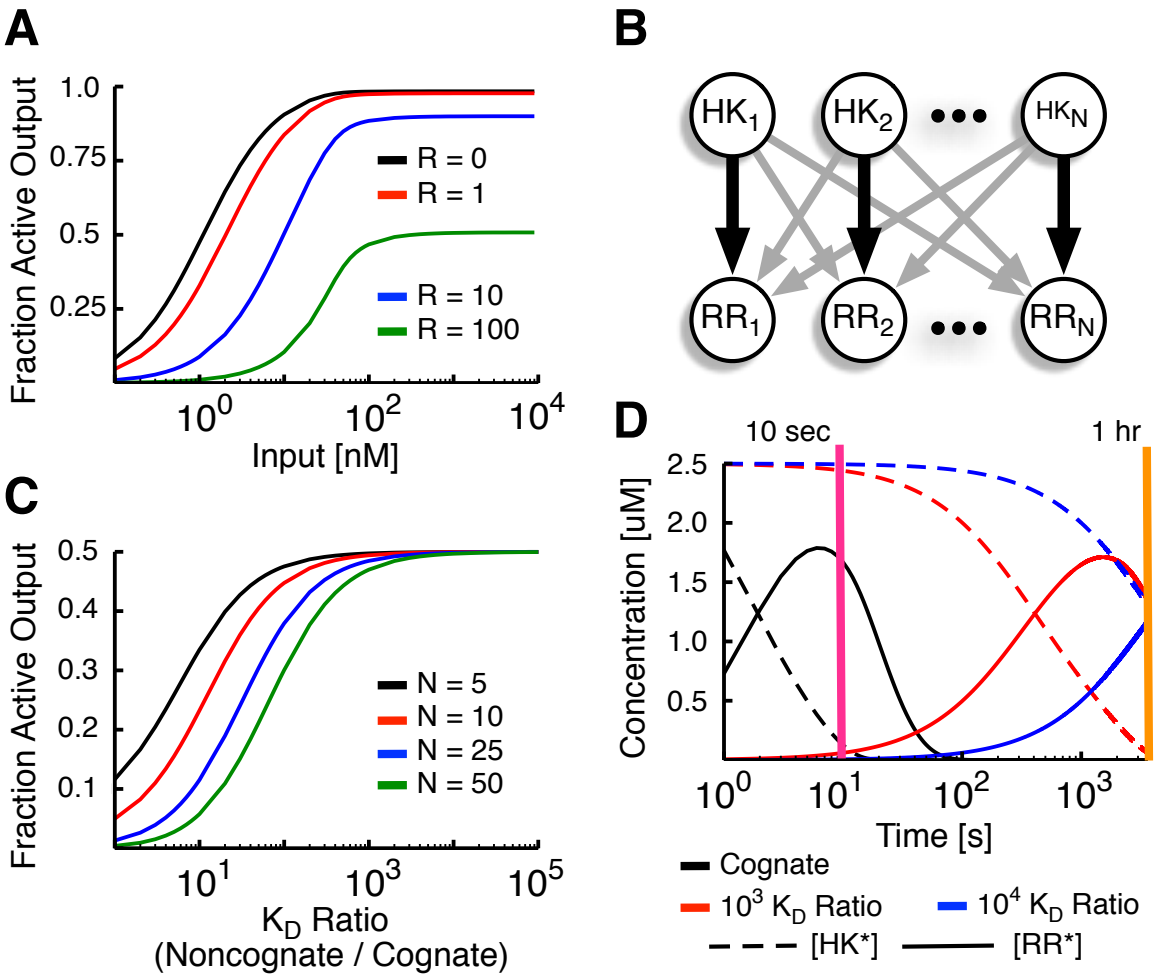

Figure 4.2: Effects of competition on TCS signaling. (*A*) Fraction of active output as a function of input concentration in response to competition between the cognate RR_1 and noncognate RR_2. The ratio $R \equiv [RR_2]_0/[RR_1]_0$ is varied as indicated. (*B*) Diagram of a TCS network with N HKs and N RRs. Each HK_i interacts with its cognate RR_i with $K_{D,C}$ (black arrows) and with noncognate RR_j with $K_{D,NC}$ (gray arrows). (*C*) Fraction of active output as a function of K_D ratio = $K_{D,C}/K_{D,NC}$ in TCS networks of varying size N. The input concentration was set at a concentration that produces 50% phosphorylation for an isolated HK-RR pair. (*D*) Concentration of HK^* (dashed lines) and RR^* (solid lines) as a function of time for a cognate substrate and two noncognate substrates with K_D ratios of 10^3 and 10^4. These models start with 2.5 μM HK^* and 2.5 μM RR, exactly replicating the in vitro experiments of Skerker, *et al.* [28]. The two time points investigated experimentally in that work are highlighted, 10 *s* (pink vertical line) and 1 *h* (orange vertical line).

with each RR; in principle,e very pair in this case has an independent association affinity (i.e., K_D). To simplify the problem, we assigned every cognate pair in the system ($HK_i - RR_i$, e.g., HK_1 interacting with RR_1) the same affinity $K_{D,C}$, and every noncognate pair ($HK_i - RR_j$, $i \neq j$, e.g., HK_1 interacting with RR_2) the same affinity $K_{D,NC}$ (Fig. 4.2*B*). We fixed the cognate interaction to the value observed experimentally ($K_{D,C} \approx 1$ μM) [125]. We then varied the ratio between noncognate and cognate K_D's (K_D ratio $\equiv K_{D,NC}/K_{D,C}$) in networks of various sizes N and measured the output activity of RR_1 in response to the activation of HK_1 (Fig. 4.2*C*). We find that output activity is heavily attenuated for all systems when the noncognate K_D's are relatively strong. However, when the noncognate K_D's are weaker than the cognate's by approximately three to four orders of magnitude, the activity of the single active pathway is essentially unaffected for $N = 5$ to $N = 50$.

To determine whether K_D ratios in this range provide "kinetic preferencing" similar to that observed by [28], we replicated their in vitro experiment using our model. This involved mixing either a cognate or noncognate RR with a fully phosphorylated HK at equal concentrations. When the HK acts on a cognate substrate, the phosphorylation of the RR peaks at 10 s, and after 1 h both the HK and RR are completely dephosphorylated. In contrast, a noncognate substrate with a K_D ratio of either 10^3 or 10^4 exhibits no phosphorylation at 10 s, but considerable response after 1 h (Fig. 4.2*D*), directly recapitulating the findings of [28]. A K_D ratio of 10^4 also gives cognate catalytic efficiencies (k_{cat}/K_m) that are 10^4 higher than noncognate efficiencies, consistent with other experimental findings [6]. Our results thus indicate that the observed kinetic preference of HKs for their cognate substrates can be explained simply by the need to maintain cognate responses (Fig. 4.2*C*) in the presence of competing substrates, rather than an explicit pressure to prevent misregulation of noncognate targets [119].

4.3.3 Evolutionary trajectories

New TCS pathways can arise through the duplication and divergence of existing HK-RR pairs [31, 121]. The duplication event itself produces two HK-RR pairs that are identical (Fig. 4.3*A*,

steps 0 to 1). This effectively increases both the total concentration of the substrate and the concentration of the HK, both of which can decrease the response of the "parent" signaling pathway (Fig. 4.2 and Appendix C). Because such decreases could strongly affect the fitness of cells in which the duplication occurs [119], the unavoidable crosstalk that occurs immediately postduplication could present a barrier to the evolution of new TC signaling pathways. Subsequent evolutionary events, such as the evolution of one duplicate RR that cannot activate the original output genes but still competes with the original RR for phosphorylation by the HKs, could easily exacerbate this problem (Fig. 4.2).

We thus determined whether there were any "evolutionary trajectories" that could minimize the effect of crosstalk on the parent signaling pathway. To do this,w e developed a simple model of the evolution of HK-RR pairs postduplication. In this model, we defined two types of evolutionary steps: the removal of an interaction, meaning that the kinetic parameters of the interaction are set so weak that binding of the two molecules becomes very unlikely, and the addition of an interaction, meaning that the kinetic parameters for the binding of two molecules are made stronger. There are thus six specific events that can occur in our evolutionary trajectories: (A) removing the $HK_2 - RR_1$ interaction, (B) removing the $HK_1 - RR_2$ interaction, (C) adding the $I_2 - HK_2$ interaction, (D) removing the $I_1 - HK_2$ interaction, (E) removing the $RR_2 - O_1$ interaction, and (F) adding the $RR_2 - O_2$ interaction. This provides a model with 64 possible states, depending upon the existence of these size interactions, and 720 possible trajectories (e.g. A, B, C, D, E, F or E, A, D, F, C, B). An example of one such trajectory is diagrammed in Fig. 4.3*A*. Each trajectory was then analyzed at each step for the activation of both outputs in the presence of either input. The neutrality of the trajectories was measured based upon a single criterion: having minimal impact on parental signaling, which we defined using the total concentration of active O_1 in the presence of saturation concentrations of I_1, summed across all of the "steps" in the trajectory.

We obtained 24 "near-neutral" trajectories that minimize impact on parental signaling equally well across all steps (Fig. 4.3*B*); the example trajectory in Fig. 4.3*A* is a member of that set. In all of these trajectories, the crosstalk interactions between the HKs and RRs are removed before

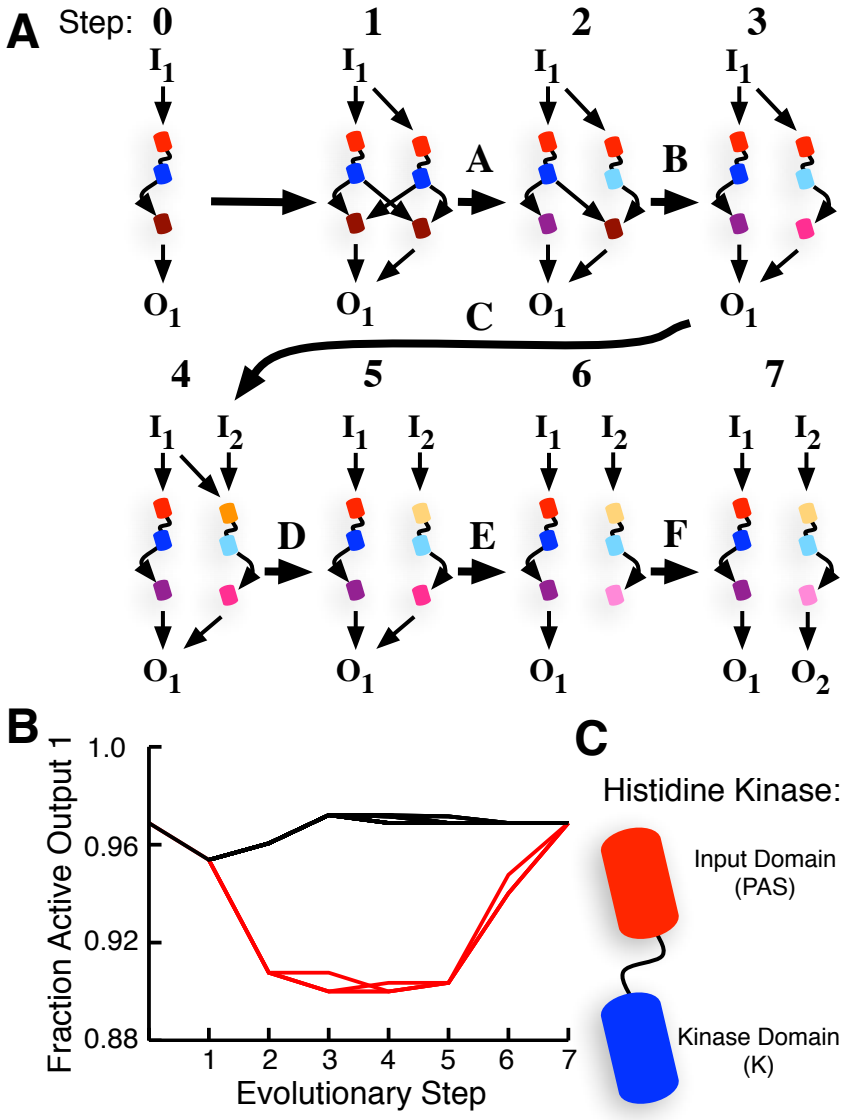

Figure 4.3: Evolutionary trajectories (*A*) An example of an evolutionary trajectory starting with a single TCS pathway in which the input I_1 activates HK_1, HK_1 phosphorylates RR_1, and RR_1 activates the output O_1. The HK-RR pair is duplicated in the first step (step $0 \rightarrow 1$), introducing crosstalk. The new HK-RR pair is modified through a series of coarse-grained events. Each step corresponds to a discrete change in the interaction capabilities of the molecules in question (events A-F described in the text). Any alternative ordering of these events constitutes a unique evolutionary trajectory. (*B*) Fraction of active output O_1 in response to saturating I_1 at each evolutionary step for the 24 trajectories that displayed the least impact on parental signaling (black) and the 4 trajectories that displayed the largest impact on parental signaling (red). Multiple trajectories can exhibit the same trends in parental signaling; hence, there are only a few visible curves. (*C*) Diagram of an HK containing two domains: an input domain *I* (PAS domain) and the kinase domain *K*.

HK_2 and RR_2 lose their capacity to interact with I_1 and O_1. This prevents inactive HK_2 from acting as a phosphatase for RR_1, and avoids reductions in O_1 activation owing to competition between RR_1 and RR_2 for phosphorylation by HK_1 (Fig. 4.2*A*). The red curves in Fig. 4.3*B* represent the four trajectories with the maximal total impact on parental signaling. These trajectories all exhibit the opposite order of events: in those cases, input/output functionality is always altered before the HK-RR crosstalk is removed.

4.3.4 Evidence for near-neutral trajectories

HK proteins generally contain a distinct "kinase" (K) domain, which interacts with the RR and is involved in the phosphotransfer reaction, as well as an "input" domain (I) that recognizes external signals and modulates HK function (Figs. 4.1*B* and 4.3*C*) [2, 9, 31, 121, 126]. Our model predicts a pressure to eliminate crosstalk relatively early in the evolutionary trajectory of a given sequence pair, before changes occur in the input domain. In evolutionary terms, this pressure would manifest itself as a set of amino acid changes in the HK-RR interaction interfaces of the duplicate pairs to insulate the two pathways from one another [6, 9, 30, 118, 119].

To test these predictions, we obtained the amino acid and DNA sequences of HKs from bacterial genomes in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [20]. Using available Pfam annotations [122], we restricted our analysis to sequences that contain a PAS domain, because this is the most common and well-studied input domain for HK proteins [121, 126]. We retained only those genomes where we could identify five or more such sequences, resulting in a total of 352 bacterial genomes. To identify putative recent duplication events, we performed multiple sequence alignments of the K domains from each genome separately and focused only on those pairs that were nearest neighbors in the phylogenetic trees obtained from those alignments.

Although duplication and divergence are common in the evolution of HKs, new pathways can also enter a lineage through horizontal gene transfer (HGT) [31, 121]. To remove HGT pairs from our analysis, we followed the approach of Alm, *et al.* [121] exactly, constructing a "phylogenetic profile" for each HK gene in our dataset based on its presence or absence across the phylogenetic

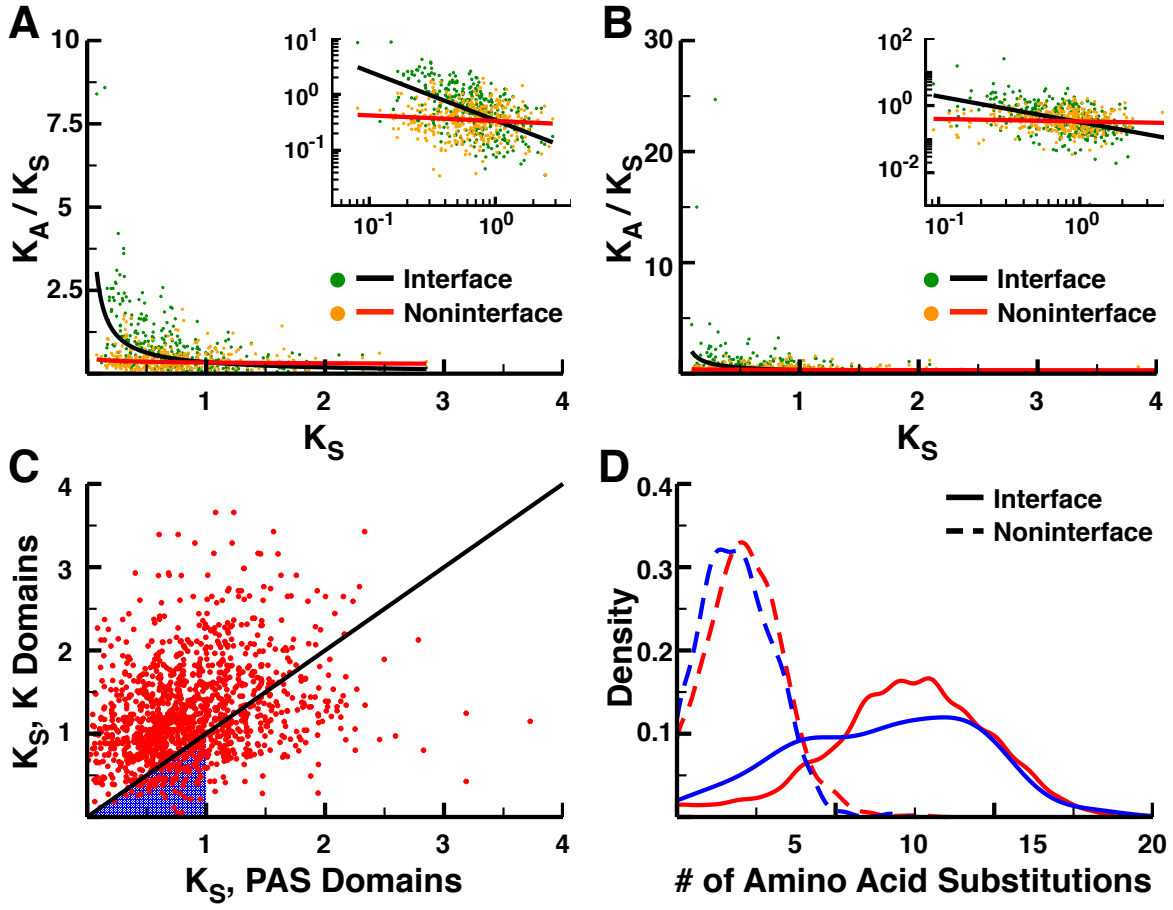

Figure 4.4: Sequence analysis. (*A*) The K_A/K_S values as a function of K_S for non-HGT HK sequence pairs for both the K domain interface residues (green circles) and noninterface residues (orange circles). The black and red lines correspond to power-law regressions of the interface and noninterface data, respectively. (Inset) the same data and fits, plotted on a log-log scale. (*B*) A plot similar to that in *A*, but for the interface and noninterface residues of RR proteins. (*C*) the K_S value for each HK domain pair is plotted against the K_S value for the corresponding PAS domains. Of the 1,300 poitns in this plot, 951 are above the diagonal (the black line, $p < 2 \times 10^{-16}$). (*D*) A plot of the distribution of substitution rates for all of the K domains in *C* (red) and just those K domains from very recent duplications ($K_S < 1$) where the PAS domain is younger (the blue triangle in *C* corresponds to the points used to make the blue line). The number of amino acid substitutions in the interface positions (solid lines) is compared with the umber obtained from random subsets of noninterface positions of the same size (dashed lines).

tree of our bacterial genomes. Of the 2,243 closely related nearest-neighbor pairs we identified, 342 of them (15%) represented recent HGT events. We thus obtained a total of 1,901 pairs that represented bona fide duplication events, at least according to this analysis. Further details regarding the sequences we obtained and the HGT analysis can be found in Materials and Methods and Appendix C.

We used this data to calculate the rate of synonymous substitutions (K_S) and the rate of nonsynonymous substitutions (K_A) for our sequences [127]. We included the sequence of Spo0B in our K domain alignments, using the available cocrystal structure between Spo0B and Spo0F (its RR) to determine which residues in each HK sequence were likely to participate in this interface [9, 128]. Using the alignment for each non-HGT pair, we calculated K_A and K_S values for the interfacial residues of the K domain and the noninterfacial residues of the K domain.

In Fig. 4.4*A*, we plot the value of K_A/K_S as a function of K_S (a rough estimator of time since duplication) for non-HGT sequence pairs based on all residues in either the K domain interface of the noninterface region. We found that the strength of selection on these subsets of residues was quite different: for one, the average K_A/K_S in the interfacial residues is higher overall (Appendix C, Fig. C.9, $p = 4.73 \times 10^{-9}$). We also found a strong power-law dependence of K_A/K_S on K_S for the interface, whereas noninterface residues showed a statistically distinct and much weaker dependence ($p < 2 \times 10^{-16}$, Appendix C, section C.5.3).

To test whether the size of the subset of residues considered might influence the calculation of K_A and K_S, we generated random subsets of noninterface residues with the same total number of residues as the interface. We also used the Spo0B structure to generate similar random subsets of noninterface surface residues, to control for the fact that surface residues (such as those on the HK/RR interface) might experience relaxed evolutionary pressures. In both cases, the trends were the same as those in Fig. 4.4*A* (Appendix C, Fig. C.6*A* and *B*). Using a second available HK/RR structure to determine the interface residues [HK853/RR468 from *Thermotoga maritima* [27]] also gives similar results (Appendix C, Fig. C.7). Finally, the raw substitution rates (i.e., the total number of amino acid changes between two sequences) shows much higher values for

interface positions compared with other positions in the sequences, regardless of whether these positions are on the surface or not (Appendix C, Fig. C.10). The difference in substitution rates can be readily seen in an example alignment for a recently diverged pair of K domains from the bacterium *Halococcus turkmenicus* (Appendix C, Fig. C.11). Using a similar analysis for RR proteins, we found essentially the same trends when comparing interface to noninterface residues for those proteins (Fig. 4.4*B* and Appendix C, Fig. C.8*A* and *B*). Overall, these findings indicate that the interface residues of both the HK and RR proteins tend to diversify after duplication to prevent crosstalk, consistent with our predictions (Fig. 4.3).

We also considered the evolution of input functionality in HK proteins. In our alignments, we found only 67 cases out of the 2,243 nearest-neighbors in the K domain alignment where the PAS domains for those two proteins were also nearest neighbors in the PAS domain alignment for that genome. In other words, we found that PAS domains tend to display extensive domain swapping, where new input functionality evolves not through divergence of the ancestral input domain, but rather the replacement of the original function through wholesale introduction of the input domain from another, unrelated protein. This is consistent with earlier findings on PAS domain evolution in HKs [121].

Because the evolution of input functionality is dominated by domain swapping, we could not perform a robust K_A/K_S analysis similar to that in Fig. 4.4*A* and *B*. Instead, we focused on understanding the timing of the domain-swapping event relative to the duplication of the HK gene. There are two possible scenarios in this case: in scenario A, an HK gene is duplicated and subsequently picks up a "new" PAS domain from some other protein in the genome. In scenario B, a protein with a PAS domain is duplicated, and later picks up a new K domain through domain swapping. Our model predicts that scenario A should be more common in HK evolution, because input changes should occur relatively later in the evolutionary trajectory (Fig. 4.3).

To test this prediction, we took each of our domain-swapped non-HGT HK pairs and compared the K_S of the closest PAS domain. Of the 1,300 cases for which we could obtain the relevant K_S values, 951 of them had a larger K_S value for the K domain than for the PAS domain (Fig. 4.4*C*,

$p < 2 \times 10^{-16}$), as our model predicted. This statistical bias is present if we consider only those vases where the PAS domain that is swapped originates only from other HK proteins, or only from non-HK proteins ($p < 2 \times 10^{-16}$ in both cases, Appendix C, Fig. C.9). Even in cases of very recent duplications, where scenario B seems more likely (i.e., the blue triangle in Fig. 4.4*C*, with K_S for both domains < 1), we see significant pressure to mutate interface residues. In particular, the average number of substitutions in the interface for those sequence pairs is approximately eight, similar to that observed for all pairs in the dataset (Fig. 4.4*D*). The average substitution rate in this case is much larger than that observed for random noninterface subsets of the same size ($p < 10^{-5}$, permutation test). This indicates a near-universal pressure to diversify the interface residues of newly evolved HK/RR pairs.

4.4 Discussion

The results described above indicate that the vast global differences in topology between eukaryotic and bacterial signaling networks are likely the result of differences in the atomic "motifs" from which the networks themselves are constructed. In particular, the kinase-phosphatase pairs that are typically found in eukaryotic networks become more ultrasensitive as they become more saturated, a behavior that allows these loops to couple the responses of multiple downstream targets in interesting and potentially adaptive ways (Fig. 4.1*C*) [26, 72]. In contrast, the two-component architecture of bacterial signaling motifs makes them inherently less efficient as they become saturated, ultimately driving down total system response as competitive substrates are added (Fig. 4.1*D*). This behavior likely underlies the fitness cost of crosstalk observed in vivo, resulting in a natural evolutionary pressure to maintain isolated cognate signaling pathways [119]. Indeed, our models indicate that a requirement to maintain cognate responses is sufficient to obtain the degree of kinetic preference that HKs show for their substrates (Fig. 4.2). [28].

Although our models indicate that crosstalk in TCS generally decreases response, this does not imply that such systems absolutely cannot tolerate the presence of more than one interaction

partner. Indeed, there are known examples of HKs that act efficiently on more than one RR (e.g., the bacterial chemotaxis pathway) [129]. Although introducing crosstalk does decrease response, the system can compensate by increasing the total expression level of that particular RR to maintain a particular concentration of active RR^* (Appendix C) [32]. Of course, such an increase comes with its own fitness cost: the bacterium must invest more energy in protein synthesis to obtain the same level of signaling performance. In some cases, the phenotypic benefits of crosstalk outweigh this cost, resulting in HKs with more than one target. As the number of targets increases, however, the cost of maintaining the response becomes larger (Fig. 4.2). This is likely the reason that even the few bifunctional HKs that have more than one target rarely act on more than two or three downstream RRs [20, 129]. Monofunctional HKs, however, should act more like kinases in GK loops (Fig. 4.1), and so proteins like the chemotaxis kinases may experience a considerably relaxed constraint against evolving crosstalk.

The evolution of new signaling pathways in bacteria often involves the duplication and subsequent divergence of an existing TCS (Fig. 4.3*A*) [31]. Our findings indicate that the impact of crosstalk on HK signaling likely shapes the evolutionary landscapes of these duplicate pairs. Specifically, the fitness costs of crosstalk generate significant evolutionary pressure that result in rapid diversification of the HK-RR interface, insulating the protein interactions and allowing the subsequent evolution of new input and output functionalities (Figs. 4.3 and 4.4). It is currently possible to engineer both HK and PAS domain sequences to introduce a wide variety of HK-RR and HK-I interactions. It would thus be straightforward to create a number of the intermediate evolutionary "states" considered by our model (e.g., Fig. 4.3*A*) and assess their relative fitness costs in vivo. One such case has already been investigated experimentally [119]; the investigation of systems with related topologies would provide detailed tests of our predictions. The combination of these experimental efforts with more detailed phylogenetic analyses of recent duplication events [121] would ultimately result in a definitive characterization of the evolutionary trajectories of new TCS pathways.

The reliance of bacterial signaling systems on only two components results in signaling dy-

namics that cannot easily admit competitive interactions. Our work indicates that this inherent feature of TCS dynamics underlies a diverse array of observations, including kinetic preferencing [28], the evolution of protein interaction interfaces (Fig. 4.4) [6, 8, 9, 28–31, 118, 119], and the deleterious effects of crosstalk in vivo [119]. The constraint against crosstalk may limit the types of information processing available to bacterial signaling networks, with more involved computations occurring at the level of the complex gene regulatory networks downstream of RRs [130, 131].

Chapter 5

Crosstalk and the Evolvability of Intracellular Communication

5.1 Introduction

Signaling networks allow cells to process information from their environment and respond in appropriate ways to input signals. These networks are generally constructed from a set of interacting proteins; changes in the activity of these proteins across the network transmits the signal from the cell membrane to downstream elements, ultimately resulting (on average) in a particular phenotypic response (e.g., proliferation, apoptosis, or differentiation). Traditionally, these networks have been organized into a set of "canonical pathways" corresponding to sets of proteins that are involved in the transmission of a specific signal [13, 16, 19–21, 132–134]. For example, the human signaling network includes pathways that are activated by Insulin-like Growth Factor-I (IGF-I), Wnt, or apoptotic signals [21, 132–134]. Although they are often studied separately, these pathways can demonstrate a high degree of "crosstalk," where proteins that are shared between two pathways cause one pathway's activity to be modulated by the activity of another [12, 13, 16, 17, 40, 72].

The degree of crosstalk present in signaling networks varies widely across evolution. For

instance, bacterial two-component signaling (TCS) networks possess little crosstalk, with most histidine kinases (HKs) acting on a single target (Fig. 5.1*A*) [6, 9, 28]. We recently demonstrated that this lack of crosstalk is likely a result of the fact that the histidine kinases that make up these networks are generally bifunctional, acting as both kinase and phosphatase for their substrates [2, 32, 135]. In contrast, metazoan networks display incredible levels of crosstalk (Fig. 5.1*B*): Kirouac et al. recently showed that there are generally more connections between canonical pathways than within them [19]. Crosstalk in metazoan networks is thus so extreme that the individual pathways generally can no longer be discerned once they are combined into a single network (Fig. 5.1*B*).

While the kinases and phosphatases typical of metazoan signaling clearly do not share the same enzymatic constraints as bacterial TCS systems [72, 135], the fact that they can display crosstalk does not explain why it is so extensive. In this work, we explore the hypothesis that the extensive crosstalk present in metazoan networks has evolved, at least in part, due to the constraints multicellularity has placed on intracellular communication. In particular, metazoans have multiple different cell types, many of which need to react differently to the same stimulus (Fig. 5.1*C*). For example, during wound healing endothelial cells construct new blood vessels, fibroblasts establish the new extracellular matrix, and epithelial cells proliferate and migrate to close the skin [136–138]. If metazoan cells contained TCS-like networks (i.e. networks with essentially no crosstalk, Fig. 5.1*A*), they would need to evolve a new signaling molecule (e.g. a new cytokine) and cognate receptor for each combination of cell type/response they needed to control separately. In organisms as complex as mammals, this approach would likely require thousands of unique cytokines, each binding specifically to one of thousands of unique receptors. A TCS-like architecture would also present a major barrier to the evolution of new cell types, since each new cell type would require the evolution of a unique complement of signals and receptors (i.e. its own "signaling channels").

We posited that the extensive crosstalk present in metazoan networks, combined with differences in the expression of various nodes in the network in different cell types, might allow those cell types to all respond differently to precisely the same set of signals. This would allow metazoans to encode a wide diversity of responses using a relatively small number of cytokines and

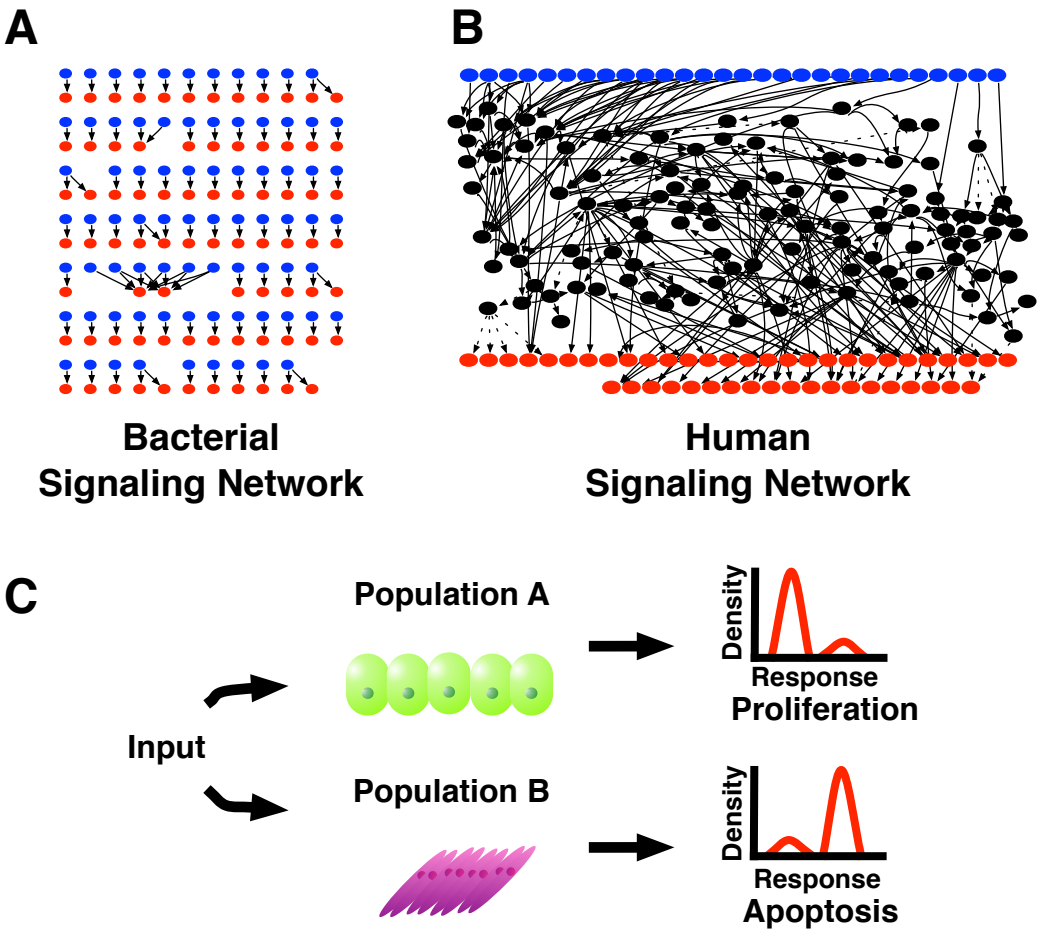

Figure 5.1: Bacterial TCS versus human signaling networks. (*A*) Diagram of the bacterial TCS network. Blue nodes represent the inputs, sensor HK's, and red nodes represent the outputs, RR's. Note the highly linear nature of the TCS network: the majority of inputs only point to one, or at times, two RR's. (*B*) Diagram of the human signaling network. This representation draws from six of the 29 canonical pathways from the KEGG pathways database [20]. Blue nodes are inputs, which are identified by searching for the keyword 'Receptor' in the UniProt entries of the genes associated with each node [89]. The red nodes are outputs, which are similarly identified by searching for the keyword 'Transcription regulation'. Note that even with six canonical pathways the network is very interconnected and is difficult to globally comprehend. (*C*) Two human cell populations from different cell lines are exposed to the same input. However, while the majority of Population A proliferates in response to the input, the majority of Population B undergoes apoptosis.

receptors. To test this hypothesis, we first created a simple Boolean network with only two inputs and two outputs, and evolved these networks to maximize the number of unique input/output maps the network could exhibit. We found that the evolved networks, which exhibited significant crosstalk and complexity, could generate hundreds of unique input/output maps with only two inputs and two outputs.

To test this idea in an existing biological system, we generated a large-scale Boolean network model of the signaling network inside human cells by combining 29 signaling pathways curated in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [20]. Using available data on protein expression [139], we found that the architecture of the signaling network changes in different tissues. In the complete signaling network, every input is able to directly signal to each of the transcription regulating outputs. However, when nodes are removed based upon expression data in various tissues, none of the inputs are able to directly signal to every output. Despite this, the cell-specific subnetworks retain a high degree of input-output interconnectivity when compared to subnetworks in which nodes are expressed at random, indicating that the expression of the signaling proteins in different tissues have specifically evolved to maintain extensive crosstalk. As a result of this connectivity, the cell-specific networks demonstrate a much greater diversity in responses to stimuli when compared to either cells with random expression vectors or cells with TCS-like networks that have no crosstalk.We found that the tissue-specific networks generally respond quite differently to the inhibition of individual proteins. These results imply that the complex interplay between network topology and gene expression that allows different cell types to respond differently to precisely the same signals has important consequences for the development of drugs that target signaling networks [140–142].

5.2 Materials and Methods

The evolvable networks are a set of Boolean networks that are randomly altered through one of three possible modifications: (1) adding an edge, (2) flipping an edge from activating to inhibiting

or vice-versa, or (3) adding an intermediate nodes to randomly connect two existing nodes. Following a modification, the network is run with one, both, or neither of the inputs active using a custom synchronous Boolean simulator for 100 steps. We then combine the final activity of the outputs to form an 8 digit binary string (i.e., '00101101'), which we term the 'I/O map'. If the network has one or more intermediates, we then run the set of simulations to obtain an I/O map for every possible expression vector (i.e., a network with two intermediate nodes has four expression vectors: both expressed, one expressed, the other expressed, and neither expressed). Acceptance of a modification depends upon the number of unique I/O maps across all expression vectors (see Appendix D for more information).

The complete KEGG signaling network was constructed from KGML files from 29 canonical pathways [20]. These pathways were initially expanded so that each gene or compound in an entry existed as its own node. This also expanded the number of edges from each pathway so that each node from an entry had its own set of edges. For example, if entry 2, which includes two genes, activates entry 3, which includes four genes, then this edge was expanded to include 8 different gene pairs. Once the full network was built from the different pathways, nodes that are involved in the same set of edges were collapsed into a single node.

We then obtained the expression of each of the genes in the network in 84 tissue types from the Human Protein Atlas [139]. The expression of any particular node is dependent upon the expression of each of its associated genes: if any gene included in the node is expressed at any level within a tissue, then the node is expressed in that tissue.

To obtain an I/O map distance between two subnetworks for a particular set of active input vectors, we ran each subnetwork for 10000 steps in a custom synchronous Boolean simulator. The I/O map in this case is the set of the average activity of each output in the last 1000 steps of the simulation for each active input vector. The result is a matrix where each row contains 67 values from 0 to 1 for the average activity of each output in response to the activation of a set of inputs. The individual elements from the matrices of the two subnetworks were then compared to obtain the distance: if the elements did not match, the distance was increased by 1 (see Appendix D for

more information).

5.3 Results

5.3.1 Crosstalk and expression provide a diversity of cellular responses

The decisions cells make, such as chemotaxis, differentiation, proliferation, apoptosis, etc., are determined by the set of inputs stimulating the signaling network and influenced by crosstalk between various canonical pathways [40]. Depending on the expression of network components, a more complex system such as the human signaling network would be able to change how inputs like EGF or NGF may influence the activity of outputs, such as Akt or Erk phosphorylation. The study by Chen, et al, illustrates the response map as a phase diagram, in which the magnitude Akt and Erk phosphorylation can determine cell fate decisions between quiescence, differentiation or proliferation [143]. Changes in inputs or the network affect the activity of Akt and Erk and can yield different phenotypic responses.

To demonstrate how network complexity can regulate the cellular decision-making process, we developed an evolvable Boolean network model that begins with a simple network consisting of two inputs that each activates their a single, cognate output (see SI Appendix, section 2). At each evolutionary step, the model can perform one of three possible modifications: (1), add an edge, (2) flip the sign of an edge, or (3) add an intermediate node (Fig. 5.2*A*). New networks are evaluated according to the number of unique input-output maps they can generate, as described below; if that number increases, the modified network is "kept" by the algorithm and used for the next set of modifications. If that number does not increase, the new network is discarded and the algorithm attempts another random modification of the previous topology.

To determine the variety of responses any given network can produce, we first generated the set of possible "expression vectors" for that network. Each expression vector represents a unique pattern of presence or absence for each of the intermediate notes in the network (i.e., a network with two intermediate nodes has four expression vectors: ‘00’, in which both are absent, ‘01’,

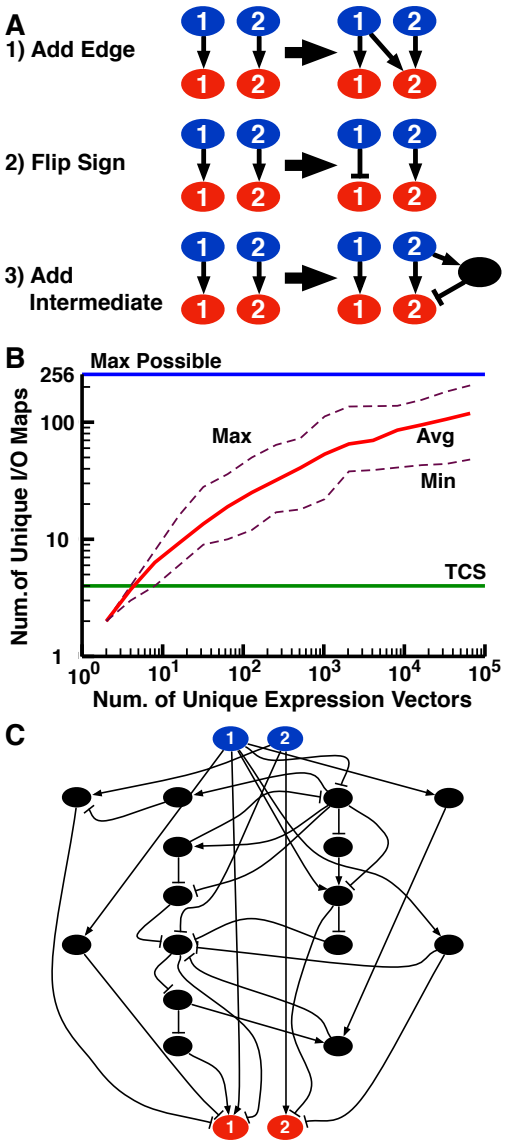

Figure 5.2: Evolvable Boolean signaling networks. (*A*) Diagram of the three possible modifications to the network. When an edge is added, the new edge connects one random node to another, and may be either activating or inhibitory. When an edge's sign is flipped, a randomly chosen edge is switched from being activating to being inhibitory or vice-versa. When an intermediate is added, the intermediate connects one random node to another. The sign of the edges to and from the new node are chosen randomly. (*B*) The number of unique I/O maps as a function of the number of unique expression vectors. A network has $2N$ unique expression vectors, where N is the number of intermediate nodes in the network. For networks with 2 inputs and 2 outputs, there are a maximum of 256 unique I/O maps possible. TCS-like networks with two intermediate nodes, each connecting a single input to a single output, has a maximum of 4 unique I/O maps, dependent upon the expression of intermediates. (*C*) An example of an evolved signaling network with 16 intermediate nodes (65536 unique expression vectors) and 32 edges connected the two inputs to the two outputs. This network generates 89 unique I/O maps, depending upon the expression of its intermediates.

'10', and '11', in which both are present). For each expression vector, we ran a Boolean network simulation with no inputs active, either input active and both active and measure the activity of the outputs at steady state. This produces an I/O map, a string of 8 binary digits that represent the activity of both outputs in response to the four combinations of active inputs (i.e., the input activities 00-01-10-11 might produce the I/O map 01-10-11-00). From these simulations, we can determine the number of unique I/O maps the network is able to generate across all expression vectors.

We began this evolutionary algorithm with a TCS-like network where each pathway includes an input activating an intermediate node, which then activates an output. The TCS-like network has only four unique expression vectors, each of which produces a unique I/O map. However, as the model evolves and includes additional intermediate nodes, the network generates a larger number of I/O maps; networks with only 16 intermediate nodes are able to produce up to 200 I/O maps (Fig. 5.2*B*). For example, the network diagrammed in Figure 5.2*C* generates 89 distinct I/O maps, depending upon the expression of its intermediates.

In networks with large numbers of I/O maps, the influence of the inputs on output activity is highly sensitive to the expression of signaling intermediates. The resulting networks also bear a fairly striking resemblance to the types of networks found in metazoan cells (Fig. 5.1*B*), exhibiting extensive crosstalk and complexity. This architecture allows for a very diverse array of responses depending on which nodes are present in the network. Even in this simple model, different cell types expressing different nodes in the network could readily exhibit widely different responses (or undertake different cell fate decisions) to the same two input signals [143].

5.3.2 Specific expression patterns in human tissues are selected to generate signaling diversity

In order to characterize the interconnectedness of the human signaling network and understand how this architecture and differential gene expression might affect responses to signals, we compiled a large Boolean signaling network by combining the contents of 29 signaling pathways from

KEGG, resulting in a network with 735 nodes and 2211 edges (see the SI Appendix for details). Using UniProt we identified 76 input nodes (keyword: 'Receptor') and 67 output nodes (keyword: 'Transcription regulation') [89]. To demonstrate the relative complexity of this human signaling network, we drew a directed graph representing only a fraction of the total network (Fig. 5.1*B*). The resulting diagram is quite elaborate, with numerous edges connected nodes from different sections of the system. Within the complete network, every output is in the forward connected component of each input (e.g., the activation of any input would have some effect on each output), revealing a significant amount of crosstalk between these canonical pathways.

The complete network we compiled from KEGG, however, does not represent the signaling networks present within different cell types due to the differences in expression of signaling proteins. In order to account for these differences, we obtained expression data for the network from the Human Protein Atlas [139]. This dataset includes relative expression levels ('High', 'Medium', 'Low', and 'Not Detected') based upon immunohistochemistry microarray assays for 344 of the 735 nodes in our network from 84 tissues. The remaining nodes either represent non-protein signaling elements, such as ions or small molecules, or were not included in the dataset, usually due to experimental constraints (e.g. the lack of a specific antibody for that protein). From the expression data we created 84 subnetworks where a node and its associated edges were removed if the node is not detected in the respective tissue. The remaining 391 nodes are always expressed.

The complete network has a different structure when compared to the set of expressed subnetworks: while each input can directly affect each output in the complete network, each signal can only reach about 50-64 of the 67 outputs in the expressed subnetworks (Fig. 5.3*A*). Thus, while the complete network serves as a useful summary of the possible interactions among proteins in the human cell signaling network, differences in the presence or absence of various nodes clearly generates a diverse set of network topologies in various tissues.

The average number of outputs each input can reach in an expressed subnetwork strongly correlates with the fraction of nodes present in the subnetwork (Spearman's $\rho = 0.8838$) (Fig. 5.3*B*, blue). To better understand subnetwork topology, we compared the real expressed subnet-

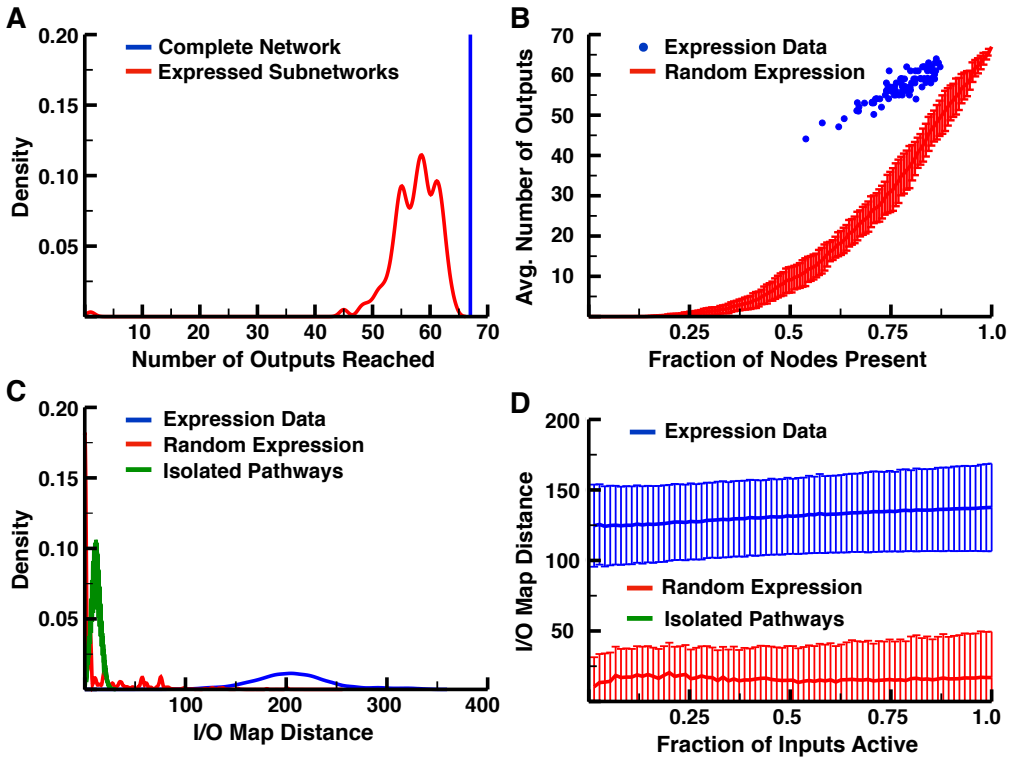

Figure 5.3: The structure and response diversity of the complete KEGG network. (*A*) The kernel density plot of the number of outputs that are downstream of each of the inputs in the complete network (blue) and each of the expressed subnetworks (red). All 67 outputs are downstream of each of the inputs in the complete network. However, most inputs have 50-64 of the outputs in their forward connected component in the expressed subnetworks. (*B*) The average number of outputs that are downstream of each of the inputs in a subnetwork versus the fraction of nodes expressed in the subnetwork. The blue dots include the connectivity of the subnetworks constructed according to the expression data from the Human Protein Atlas [139] while the red lines show the mean and standard deviation of 10 subnetworks constructed through the random expression of nodes. Note that the average number of outputs downstream of any input is much lower in the randomly expressed subnetworks, indicating that these subnetworks demonstrate less interconnectivity than those based upon expression data (minimum difference = 9.398, $p = 1.546 \times 10^{-4}$). (*C*) The kernel density plot of the I/O map distances of the expressed subnetworks (blue), randomly expressed subnetworks (red), and the TCS-like networks (green). In this panel distances are based upon the activation of each of the inputs individually. (*D*) The average and standard deviation of the I/O map distance versus the fraction of inputs active for the expressed subnetworks (blue), randomly expressed subnetworks (red), and the TCS-like networks (green). In this panel distances are based upon the activation of 50 combinations of inputs, which is why the averages are smaller than those in *C*.

works to a set of random subnetworks of varying sizes (Fig. 5.3*B*, red). For any given network size, it is clear that the expressed subnetworks found in human tissues maintain a significantly higher connectivity between inputs and outputs (minimum difference = 9.398, p = 1.546×10^{-4}). The expression of individual nodes in human tissues does not depend upon their indegree, outdegree, or their betweenness with regard to inputs to outputs, common properties that are used to quantify the "topological centrality" of nodes in a network. This suggests that the non-random connectivity we observe depends on global properties of the set of nodes that are present and absent in various human tissues (see Appendix D, Fig. D.1).

We then performed a set of Boolean network simulations on each expressed subnetwork to evaluate whether their non-random architecture had an impact on their functional properties. For any given subnetwork, we independently activated each input, running the model to steady state to obtain the I/O map. To account for possible feedback loops, the I/O map of an input is determined by the average activity of each output in the last 1000 steps of the simulation. The I/O map distance for a pair of tissues is calculated as the number of differences between the two I/O maps (see Appendix D). The higher the I/O map difference between two networks, the less similar their responses are across all possible inputs. We determined the I/O map distances for each pair for the expressed subnetworks and the randomly expressed subnetworks. The average distance for the expressed subnetworks is around 200, yet the distances for the random subnetworks is much closer to 0, indicating that the expressed subnetworks are not only more connected, but they also demonstrate a higher signal diversity (Fig. 5.3*C*) ($p < 2.2 \times 10^{-16}$). To compare the expressed subnetworks to an isolated, TCS-like network, we randomly connected input nodes to outputs so that each input only activated a single output. This network included each node from the complete KEGG network so that each input's signal propagated through about the same number of nodes before reaching the output and no intermediate node was activated by more than one input. Subnetworks of the TCS-like network were created using the expression data from the Human Protein Atlas. The I/O map distances of the TCS-like network are even lower than those from the random subnetworks, indicating cells with isolated pathways show very little diversity ($p < 2.2 \times 10^{-16}$ for

the comparison with either the actual human networks or the random subnetworks).

Although the above analysis is informative, we only activated a single input at a time, and it is unclear that the same patterns of signaling diversity would be observed when multiple inputs are activated. To test this, we created random combinations of inputs with a total of $N = 1, 2, \ldots, 75$ inputs being activated. For each value of N, we created 50 independent combinations with that total number or active inputs. We found that the difference in I/O distances between the three types of networks (i.e. actual human tissues, random subnetworks and the TCS-like network) does not depend strongly on the number of active inputs (Fig. 5.3*D*).

Our results indicate that the diversity of responses to input depends both on the complexity of the network and the targeted expression of signaling proteins. Signaling diversity does not seem to be dependent only on the number of nodes in the network as the sets of subnetworks include systems of the same size, yet the expressed subnetworks possess a higher average I/O map distance. Instead, diversity is grounded on the complex architecture of the network, with expression of signaling proteins being driven by the need to maintain a well-connected structure.

5.3.3 Differential effects of inhibitors across cell types

An interesting observation from our evolved Boolean networks (Fig. 5.2) is that despite the possibility for tens of thousands of possible expression states, these subnetworks only produce an average of around 100 unique I/O maps (Fig. 5.2*B*). Modifications of network structure in this model are accepted based upon their ability to add plasticity to the network with no inherent pressure for these networks to develop robustness. We found that the individual nodes in these networks are neither completely dispensable nor completely indispensable. The presence of any particular node may result in a subset of I/O maps while its absence may result in a different subset of maps. Previous studies have found similar behaviors in complex networks: many different 'genotypes' (e.g., RNA or amino acid sequences) may have the same 'phenotype' (e.g., folded, three-dimensional structure). However, these groups of genotypes exhibit greater evolvability, allowing the system to produce a wider variety of phenotypes with very few modifications [144–147]. These results

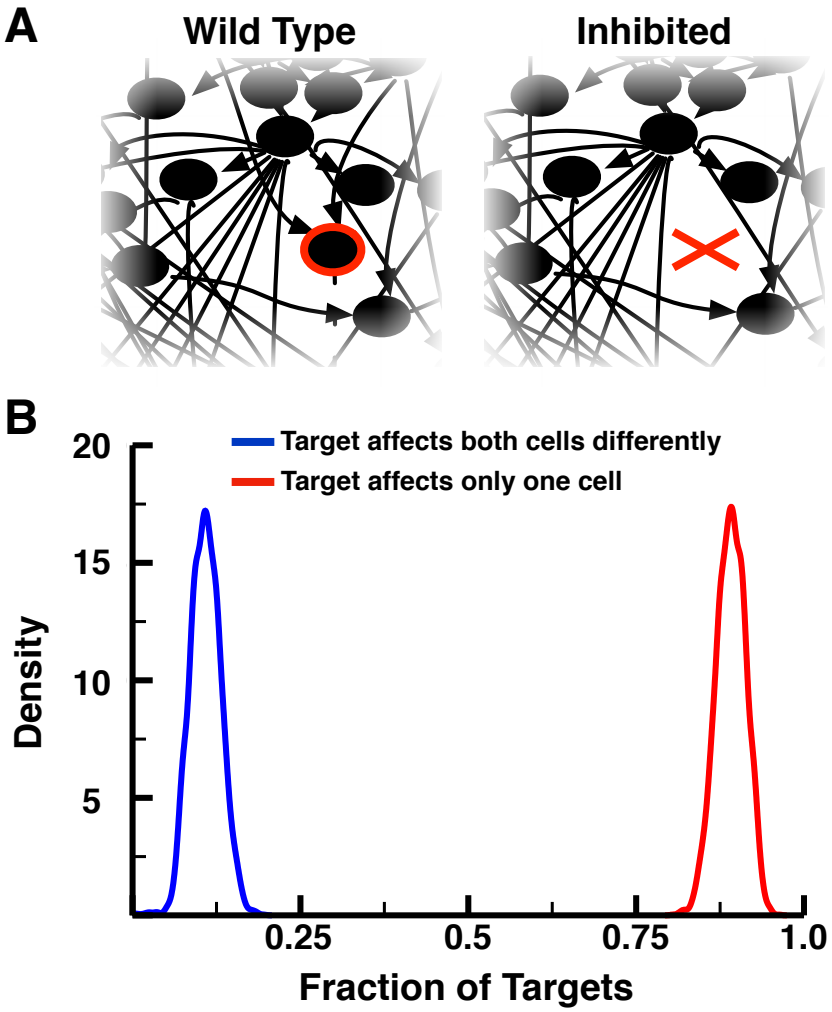

Figure 5.4: The effects of inhibitors on different cells. (*A*) Each of the nodes in complete the network were targeted for inhibition. When a node is targeted for inhibition (i.e., the red highlighted node in the wild-type panel), the node and its associated edges (including both incoming and outgoing edges) are removed from the expressed subnetworks. (*B*) The kernel density plot of the fraction of nodes whose inhibition changes the response to input in either of pair of subnetworks. Only about 12% of the targetable nodes affect both cells, and inhibiting each of these nodes affects both of the cells differently. The remaining targetable nodes affect only one cell, either because the nodes are not expressed in the other cell or because the other cell includes connections within the network that can compensate for the actions of the inhibitor.

suggest that the effect of inhibiting a single protein may alter the architecture of signaling networks of cells in different ways, depending upon the expression state of the network.

To explore the differential effects of inhibitors, every node was sequentially targeted and removed from the 84 expressed subnetworks, representing its complete inhibition in the different tissues (see Fig. 5.4*A*). We then obtained the I/O maps for the inhibited systems, which were compared to the wild-type I/O map for the respective expressed subnetwork. From this comparison we obtained an inhibitor map, where a '1' represents an increase in the average activity for an output, a '-1' represents a decrease in average activity, and a '0' signifies no change in average activity. Every node can be inhibited and result in a change in output response in at least one of the expressed subnetworks. On average, inhibiting any node in the network alters the output response in about 17 of the 84 expressed subnetworks. There are no nodes whose inhibition alters output response in all 84 subnetworks (See Appendix D, Fig. D.2).

The inhibitor maps for each targeted node across all pairs of expressed subnetworks were then analyzed to characterize the effects of inhibiting the respective node on each tissue. Of the subset of nodes whose inhibition had an effect on the I/O map of either subnetwork across all pairs, the inhibition of about 88% of the targets affected only one of the subnetworks (Fig. 5.4*B*, red). This could be because the node is not expressed in the other subnetwork, so the inhibitor had nothing to target, or because the network possesses redundancy in its structure that made the removal of that node ineffective. Inhibiting the remaining 12% of the targets affected both of the subnetworks, yet resulted in a different effect in each (Fig. 5.4*B*, blue). There are no cases where the inhibition of a single node had the same impact in two subnetworks.

The diverse effects of inhibitors on cellular responses due to the changes in network architecture arising from differential expression of components provide an explanation for the previously observed range of responses to known inhibitors. Fallahi-Sichani, et al., recently revealed a wide array of effects of various anticancer drugs against a panel of breast cell lines; drugs targeting signaling proteins often resulted a higher variation of efficacy against different cell lines compared to those targeting more ubiquitous targets such as the proteasome or DNA [148]. Quite often, cancer

cells are able to use the complexity provided by the massive amount of crosstalk in the human signaling network to compensate for drug effects [149]. Nutlin-3, for example, was designed to be a competitive inhibitor of the p53-MDM2 interaction in order to induce apoptosis or senescence [140]. However, nutlin-3 only results in reversible cell cycle arrest in most cell lines expressing wild-type p53; the induction of apoptosis or senescence is dependent upon the overexpression of MDM2 [140, 150–152]. Tissue selectivity may have consequences for the use of inhibitors and gene silencing in future signaling studies: these perturbations could have unexpectedly different effects in various cell lines. Care must be taken to consider the tissue-specific context of the signaling network to inform the interpretation of inhibitor or silencing experiments.

5.4 Discussion

RNA and protein levels have been shown to vary not only between cells from different lineages, but between cells of the same type [153]. This could potentially lead to different phenotypic responses of cells within a population to the same environmental stimuli. However, averaging the activity of cells of the same type demonstrates that these cells fluctuate around an average response [153–155]. The average response to a particular stimulus, though, differs between cells with different 'cellular barcodes', different mean RNA and protein expression [156, 157]. In this work we reveal that when protein expression is applied to the human signaling network, a series of tissue-specific subnetworks are created. These subnetworks maintain the highly complex, interconnected properties of the complete signaling network, yet the architecture is significantly different across different tissues. Moreover, the expression of particular nodes does not seem to be random, nor is it dependent upon classic individual properties of the node such as degree or betweenness. The set of expressed nodes are instead chosen to ensure interconnectivity and to produce a diversity of responses to stimuli (Fig. 5.3).

A consequence of the variable architecture of the human signaling network across different tissues is that perturbations to the network, such as small molecule inhibition, may affect the cells

in different ways. The local context of a targeted protein may be vastly different in other cells based on the expression state of its signaling neighbors. We find that targeting any node for inhibition in the network can potentially change the response of a cell to stimuli in some, but not all, tissues. However, the phenotypic consequences of the inhibitor vary between the different cell types it affects. There are no examples of an inhibitor that has the same effect in two or more tissues in our subnetwork, revealing the importance of tissue-specific context in considering drug targets.

These results provide a potential explanation for the variety of mutations across different cancers and other diseases. Protein affecting mutations (PAMs) also perturb the network through changing the activity or expression of a signaling protein. It has been observed that a PAM may not have the same phenotypic effect in different tissues, similar to how inhibitors have different effects in our tissue-specific subnetworks. For instance, mutations affecting p53 and PIK3CA, both high-confidence drivers (HCDs) of carcinogenesis, have been found in just over 10% of cancer samples. In fact, mutations in several HCDs have been discovered in only a single tumor type [158]. Our results suggest that the effects of a PAM may vary across cell lines due to differences in the expression of the signaling network. The expressed subnetwork of a particular cell line may be robust to a PAM by providing an alternative pathway to maintain input-output functionality and overall cellular decision-making.

Ultimately, our work demonstrates that the tissue-specific expression of proteins can provide significantly different, yet well-connected, signaling networks. These subnetworks on average respond differently to the same stimuli and intracellular and extracellular perturbations to the cell. The aforementioned findings described above demonstrate that the cellular context must be considered in interpreting network connectivity and guiding the development of future drugs to safeguard against unexpected nonspecific effects.

Chapter 6

Conclusion

Metazoan signaling networks exhibit a high degree of crosstalk, in which enzymes are shared between multiple canonical pathways [12–14, 16, 17, 34, 35]. Previous crosstalk studies have often focused on the sharing of kinases between pathways [12, 14, 16, 36]. Kinases and their substrates are generally well characterized. For example, Akt has been shown to phosphorylate at least 18 substrates, and the receptor tyrosine kinases of the EGF/ErbB family have over 20 identified substrates [22, 23], highlighting the potential for extensive kinase crosstalk in signaling networks.

In Chapter 2 we used mathematical modeling of simple network crosstalk motifs to understand how the competition between multiple substrates for shared enzymes may influence network behavior. We find that ultrasensitivity is transitive; the saturation of the shared enzyme by one substrate causes every substrate to react ultrasensitively to the enzyme, regardless of its ability to saturate the enzyme independently. This is because the saturation of an enzyme is cumulative. A set of proteins may collectively saturate the enzyme, causing each to respond ultrasensitively, even if none of the proteins would respond ultrasensitively on its own.

Our results also demonstrate the possibility of phosphatase-mediated crosstalk. To our knowledge, the potential for phosphatases to contribute to network crosstalk has not been previously considered [63]. However, it is unclear how pervasive phosphatase crosstalk may be in these net-

works as the targets of many phosphatases may not be fully characterized [67–69]. Additionally, we demonstrate several mechanisms by which phosphatases (or other signaling enzymes) may insulate competing substrates, thus preventing crosstalk between different parts of the network (see Chapter 3).

The behaviors of signaling networks depend not only on the architecture of the network, but also on the saturation and targeting mechanisms of the signaling enzymes. Even in cases in which the connectivity of the signaling molecules is well characterized, the intracellular concentrations and Michaelis constants of the enzymes and their substrates are often not known. In future signaling studies, acquiring these details will be crucial to our interpretation and understanding of how highly interconnected signaling and regulatory networks function and respond to stimuli.

While metazoan signaling networks display extensive crosstalk, bacterial TCS networks demonstrate a highly linear topology in which HKs often act on a single RR [4–9]. This difference in global network topology is likely due to differences in enzyme dynamics. The kinases and phosphatases typically found in eukaryotic networks are monofunctional and their substrates respond more ultrasensitively as they saturate the enzymes. As such, these enzymes can be shared between multiple substrates to produce a variety of response behaviors. In contrast, HKs are often bifunctional: they act as both kinase and phosphatase for their cognate RRs [2, 3, 27, 32, 33, 115–117]. We find that the saturation of bifunctional HKs results in the HKs being less efficient at phosphorylating their RRs (see Chapter 4). Competition between multiple RRs for the same HK would increase the saturation of the enzyme, driving down the total response of each involved pathway. As such, there would be a natural pressure to preserve the insulation between pathways found in bacterial TCS networks in order to maintain cognate signaling responses.

The pressure to insulate TCS pathways, however, suggests a barrier in the evolution of new pathways. Using a coarse-grained model of the evolution of new TCS pathways, we demonstrate that the resulting HK-RR pairs experience post-duplication pressure to rapidly insulate the protein interactions prior to evolving new input/output functionalities. In future experimental studies, it would be relatively straightforward to recreate the intermediate states considered by our model to

test our predictions of the relative fitness costs. Combined with a more detailed phylogenetic analysis of recent duplications, future studies could definitively characterize the evolutionary trajectories of new TCS pathways.

While the kinases and phosphatases typical of metazoan signaling do not share the same enzymatic constraints as the bifunctional HKs of bacterial TCS networks, the fact that they can display crosstalk does not explain why it is so extensive. In Chapter 5, we explored the hypothesis that different levels of crosstalk found in various organisms have evolved, in part, to accommodate the various functional roles these networks must provide. A Boolean network with just sixteen highly interconnected intermediate nodes would be able to provide a cell with potentially hundreds of different responses to just two inputs, depending upon the structure of the subnetwork created by the expression of different combinations of intermediate nodes. This is a result of a phenomenon we term 'compact coding': more complex networks are able to achieve a large diversity of cellular responses while maintaining a relatively low number of unique signaling molecules and receptors due to their extensive crosstalk.

In Chapter 5, we produced a series of tissue-specific cell types by applying protein expression profiles of signaling proteins from 84 tissues to the human signaling network. The subnetworks generally retain the highly interconnected properties of the complete network, yet the architecture of the subnetwork differs across various tissues. The expression of these signaling proteins is not random, nor is it dependent upon individual properties of nodes, such as degree or betweenness. Instead, the expression patterns have evolved to maintain interconnectivity and produce a diverse set of responses to stimuli. Interestingly, the variable architecture of the expressed signaling network in various cells results in different effects of network perturbations, such as protein-protein interaction inhibitors, on the phenotypic responses to stimuli. We find that inhibiting any particular signaling protein in the network may potentially change the responses of a cell to stimuli in some, but not all, tissues. Additionally, the phenotypic consequences of the inhibitor vary between the different tissues it affects.

Our findings demonstrate potential problems for the interpretation of network activity in dif-

ferent cell lines, including the responses to perturbations from drugs or cancer-causing mutations. The effects of a drug or mutation may vary across cell lines due to differences in the structure of the expressed signaling network, with tissues being potentially robust to the mutation due to the inherent redundancy of the network. Because of this property, the effects of inhibiting, mutating, or silencing a protein in one cell line may not provide the same effect in other cell lines. As such, the interpretation of network perturbations in previous, current, and future signaling studies needs to be considered within the context of the cells used in the experiments.

References

[1] Liao Y, Hung M (2010) Physiological regulation of akt activity and stability. Am J Transl Res 2: 19-42.

[2] Stock AM, Robinson VL, Goudreau PN (2000) Two-component signal transduction. Annual Review of Biochemistry 69: 183-215.

[3] Qin L, Yoshida T, Inouye M (2001) The critical role of dna in the equilibrium between ompr and phosphorylated ompr mediated by envz in escherichia coli. Proceedings of the National Academy of Sciences of the United States of America 98: 908-13.

[4] Ninfa AJ, Ninfa EG, Lupas AN, Stock A, Magasanik B, et al. (1988) Crosstalk between bacterial chemotaxis signal transduction proteins and regulators of transcription of the ntr regulon: evidence that nitrogen assimilation and chemotaxis are controlled by a common phosphotransfer mechanism. Proceedings of the National Academy of Sciences of the United States of America 85: 5492-6.

[5] Fisher SL, Kim SK, Wanner BL, Walsh CT (1996) Kinetic comparison of the specificity of the vancomycin resistance vansfor two response regulators, vanr and phob. Biochemistry 35: 4732-40.

[6] Skerker JM, Perchuk BS, Siryaporn A, Lubin EA, Ashenberg O, et al. (2008) Rewiring the specificity of two-component signal transduction systems. Cell 133: 1043-54.

[7] Fisher SL, Jiang W, Wanner BL, Walsh CT (1995) Cross-talk between the histidine protein

kinase vans and the response regulator phob. characterization and identification of a vans domain that inhibits activation of phob. The Journal of Biological Chemistry 270: 23143-9.

[8] Siryaporn A, Goulian M (2008) Cross-talk suppression between the cpxa-cpxr and envz-ompr two-component systems in e. coli. Molecular microbiology 70: 494-506.

[9] Laub MT, Goulian M (2007) Specificity in two-component signal transduction pathways. Annual review of genetics 41: 121-45.

[10] Maniati E, Bossard M, Cook N, Candido JB, Emami-Shahri N, et al. (2011) Crosstalk between canonical nf-kappab and notch signaling pathways inhibits ppar gamma expression and promotes pancreatic cancer progression in mice. Journal of Clinical Investigation 121: 4685-4699.

[11] Flick LM, Florczyk SJ (2007) Exploring the cross-talk between p38 mapk and canonical wnt signaling pathways. The Journal of Immunology 178.

[12] Javelaud D, Mauviel A (2005) Crosstalk mechanisms between the mitogen-activated protein kinase pathways and smad signaling downstream of tgf-beta: implications for carcinogenesis. Oncogene 24: 5742-5750.

[13] Junttila MR, Li SP, Westermarck J (2008) Phosphatase-mediated crosstalk between mapk signalling pathways in the regulation of cell survival. Faseb Journal 22: 954-965.

[14] Liu QH, Hofmann PA (2004) Protein phosphatase 2a-mediated cross-talk between p38 mapk and erk in apoptosis of cardiac myocytes. American Journal of Physiology-Heart and Circulatory Physiology 286: H2204-H2212.

[15] Caliceti C, Nigro P, Ferrari R (2014) Ros, notch, and wnt signaling pathways: Crosstalk between three major regulators of cardiovascular biology. BioMed Research International 2014.

[16] Danielpour D, Song K (2006) Cross-talk between igf-i and tgf-beta signaling pathways. Cytokine & growth factor reviews 17: 59-74.

[17] Hill SM (1998) Receptor crosstalk: communication through cell signaling pathways. The Anatomical record 253: 42-8.

[18] Maile LA, Clemmons DR (2002) The alphavbeta3 integrin regulates insulin-like growth factor i (igf-i) receptor phosphorylation by altering the rate of recruitment of the src-homology 2-containing phosphotyrosine phosphatase-2 to the activated igf-i receptor. Endocrinology 143: 4259-64.

[19] Kirouac D, Saez-Rodriguez J, Swantek J, Burke J, Lauffenburger D, et al. (2012) Creating and analyzing pathway and protein interaction compendia for modelling signal transduction networks. BMC systems biology 6: 29.

[20] Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M (2004) The kegg resource for deciphering the genome. Nucleic acids research 32: D277-80.

[21] Laviola L, Natalicchio A, Giorgino F (2007) The igf-i signaling pathway. Current Pharmaceutical Design 13.

[22] Brazil DP, Hemmings BA (2001) Ten years of protein kinase b signalling: a hard akt to follow. Trends in biochemical sciences 26: 657-64.

[23] MacBeath G (2002) Protein microarrays and proteomics. Nature genetics 32 Suppl: 526-32.

[24] Virshup DM, Shenolikar S (2009) From promiscuity to precision: protein phosphatases get a makeover. Molecular cell 33: 537-45.

[25] Kim SY, Ferrell J J E (2007) Substrate competition as a source of ultrasensitivity in the inactivation of wee1. Cell 128: 1133-45.

[26] Goldbeter A, Koshland J D E (1981) An amplified sensitivity arising from covalent modification in biological systems. Proceedings of the National Academy of Sciences of the United States of America 78: 6840-4.

[27] Casino P, Rubio V, Marina A (2009) Structural insight into partner specificity and phosphoryl transfer in two-component signal transduction. Cell 139: 325-36.

[28] Skerker JM, Prasol MS, Perchuk BS, Biondi EG, Laub MT (2005) Two-component signal transduction pathways regulating growth and cell cycle progression in a bacterium: a system-level analysis. PLoS biology 3: e334.

[29] Groban ES, Clarke EJ, Salis HM, Miller SM, Voigt CA (2009) Kinetic buffering of cross talk between bacterial two-component sensors. Journal of molecular biology 390: 380-93.

[30] Siryaporn A, Perchuk BS, Laub MT, Goulian M (2010) Evolving a robust signal transduction pathway from weak cross-talk. Molecular systems biology 6: 452.

[31] Capra EJ, Laub MT (2012) Evolution of two-component signal transduction systems. Annual review of microbiology 66: 325-47.

[32] Batchelor E, Goulian M (2003) Robustness and the cycle of phosphorylation and dephosphorylation in a two-component regulatory system. Proceedings of the National Academy of Sciences of the United States of America 100: 691-6.

[33] Ray JC, Igoshin OA (2010) Adaptable functionality of transcriptional feedback in bacterial two-component systems. PLoS computational biology 6: e1000676.

[34] Mayer BJ, Blinov ML, Loew LM (2009) Molecular machines or pleiomorphic ensembles: signaling complexes revisited. Journal of biology 8: 81.

[35] Thomson M, Gunawardena J (2009) Unlimited multistability in multisite phosphorylation systems. Nature 460: 274-7.

[36] Eliceiri BP (2001) Integrin and growth factor receptor crosstalk. Circulation research 89: 1104-10.

[37] Gomez-Uribe C, Verghese GC, Mirny LA (2007) Operating regimes of signaling cycles: statics, dynamics, and noise filtering. PLoS computational biology 3: e246.

[38] Behar M, Dohlman HG, Elston TC (2007) Kinetic insulation as an effective mechanism for achieving pathway specificity in intracellular signaling networks. Proceedings of the National Academy of Sciences of the United States of America 104: 16146-51.

[39] Kim Y, Paroush Z, Nairz K, Hafen E, Jimenez G, et al. (2011) Substrate-dependent control of mapk phosphorylation in vivo. Molecular systems biology 7: 467.

[40] McClean MN, Mody A, Broach JR, Ramanathan S (2007) Cross-talk and decision making in map kinase pathways. Nature genetics 39: 409-14.

[41] Belle A, Tanay A, Bitincka L, Shamir R, O'Shea EK (2006) Quantification of protein half-lives in the budding yeast proteome. Proceedings of the National Academy of Sciences of the United States of America 103: 13004-9.

[42] Bagowski CP, Besser J, Frey CR, Ferrell J J E (2003) The jnk cascade as a biochemical switch in mammalian cells: ultrasensitive and all-or-none responses. Current biology : CB 13: 315-20.

[43] Hardie DG, Salt IP, Hawley SA, Davies SP (1999) Amp-activated protein kinase: an ultrasensitive system for monitoring cellular energy charge. The Biochemical journal 338 (Pt 3): 717-22.

[44] Huang CY, Ferrell J J E (1996) Ultrasensitivity in the mitogen-activated protein kinase cascade. Proceedings of the National Academy of Sciences of the United States of America 93: 10078-83.

[45] LaPorte DC, Koshland J D E (1983) Phosphorylation of isocitrate dehydrogenase as a demonstration of enhanced sensitivity in covalent regulation. Nature 305: 286-90.

[46] Meinke MH, Bishop JS, Edstrom RD (1986) Zero-order ultrasensitivity in the regulation of glycogen phosphorylase. Proceedings of the National Academy of Sciences of the United States of America 83: 2865-8.

[47] Loog M, Morgan DO (2005) Cyclin specificity in the phosphorylation of cyclin-dependent kinase substrates. Nature 434: 104-8.

[48] Yan Z, Fedorov SA, Mumby MC, Williams RS (2000) Pr48, a novel regulatory subunit of protein phosphatase 2a, interacts with cdc6 and modulates dna replication in human cells. Molecular and cellular biology 20: 1021-9.

[49] Woodbury EL, Morgan DO (2007) Cdk and apc activities limit the spindle-stabilizing function of fin1 to anaphase. Nature cell biology 9: 106-12.

[50] Bloom J, Cristea IM, Procko AL, Lubkov V, Chait BT, et al. (2011) Global analysis of cdc14 phosphatase reveals diverse roles in mitotic processes. The Journal of Biological Chemistry 286: 5434-45.

[51] Matheos D, Metodiev M, Muller E, Stone D, Rose MD (2004) Pheromone-induced polarization is dependent on the fus3p mapk acting through the formin bni1p. The Journal of cell biology 165: 99-109.

[52] Enserink JM, Kolodner RD (2010) An overview of cdk1-controlled targets and processes. Cell division 5: 11.

[53] Zhang J, Yang PL, Gray NS (2009) Targeting cancer with small molecule kinase inhibitors. Nature reviews Cancer 9: 28-39.

[54] Hindmarsh PC (2005) The impact of hormone physiology on clinical practice with reference

to two papers spanning 40 years of endocrinology. Archives of disease in childhood 90: 1144-7.

[55] Khandelwal RL, Vandenheede JR, Krebs EG (1976) Purification, properties, and substrate specificities of phosphoprotein phosphatase(s) from rabbit liver. The Journal of Biological Chemistry 251: 4850-8.

[56] Radhakrishnan Y, Busby J W H, Shen X, Maile LA, Clemmons DR (2010) Insulin-like growth factor-i-stimulated insulin receptor substrate-1 negatively regulates src homology 2 domain-containing protein-tyrosine phosphatase substrate-1 function in vascular smooth muscle cells. The Journal of Biological Chemistry 285: 15682-95.

[57] Voet D, Voet J (2011) Biochemistry. Hoboken, NJ: John Wiley and Sons.

[58] Gunawardena J (2005) Multisite protein phosphorylation makes a good threshold but can be a poor switch. Proceedings of the National Academy of Sciences of the United States of America 102: 14617-22.

[59] Markevich NI, Hoek JB, Kholodenko BN (2004) Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades. The Journal of cell biology 164: 353-9.

[60] Ubersax JA, Woodbury EL, Quang PN, Paraz M, Blethrow JD, et al. (2003) Targets of the cyclin-dependent kinase cdk1. Nature 425: 859-64.

[61] Ptacek J, Devgan G, Michaud G, Zhu H, Zhu X, et al. (2005) Global analysis of protein phosphorylation in yeast. Nature 438: 679-84.

[62] Kaushansky A, Gordus A, Budnik BA, Lane WS, Rush J, et al. (2008) System-wide investigation of erbb4 reveals 19 sites of tyr phosphorylation that are unusually selective in their recruitment properties. Chemistry & biology 15: 808-17.

[63] Hunter T (2007) The age of crosstalk: Phosphorylation, ubiquitination, and beyond. Molecular cell 28: 730-738.

[64] Arena S, Benvenuti S, Bardelli A (2005) Genetic analysis of the kinome and phosphatome in cancer. Cellular and molecular life sciences : CMLS 62: 2092-9.

[65] Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S (2002) The protein kinase complement of the human genome. Science 298: 1912-1934.

[66] Walsh C (2006) Posttranslational Modification of Proteins: Expanding Nature's Inventory. Greenwood Village, CO: Roberts and Co. Publishers.

[67] Kholodenko BN (2000) Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. European journal of biochemistry / FEBS 267: 1583-8.

[68] Schoeberl B, Eichler-Jonsson C, Gilles ED, Muller G (2002) Computational modeling of the dynamics of the map kinase cascade activated by surface and internalized egf receptors. Nature biotechnology 20: 370-5.

[69] Shao D, Zheng W, Qiu W, Ouyang Q, Tang C (2006) Dynamic studies of scaffold-dependent mating pathway in yeast. Biophysical Journal 91: 3986-4001.

[70] Nelson D, Lehninger A, Cox M (2008) Lehninger Principles of Biochemistry. New York: W.H. Freeman.

[71] Wells JA, McClendon CL (2007) Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. Nature 450: 1001-9.

[72] Rowland MA, Fontana W, Deeds EJ (2012) Crosstalk and competition in signaling networks. Biophysical Journal 103: 2389-98.

[73] Kholodenko BN, Demin OV, Moehren G, Hoek JB (1999) Quantification of short term signaling by the epidermal growth factor receptor. The Journal of Biological Chemistry 274: 30169-81.

[74] Schoeberl B, Eichler-Jonsson C, Gilles ED, Muller G (2002) Computational modeling of the dynamics of the map kinase cascade activated by surface and internalized egf receptors. Nature biotechnology 20: 370-375.

[75] Ahmed S, Grant K, Edwards L, Rahman A, Cirit M, et al. (2014) Data-driven modeling reconciles kinetics of erk phosphorylation, localization, and activity states. Mol Syst Biol 10: 718.

[76] Eichhorn PJ, Creyghton MP, Bernards R (2009) Protein phosphatase 2a regulatory subunits and cancer. Biochimica et Biophysica Acta 1795: 1-15.

[77] Ren L, Chen X, Luechapanichkul R, Selner NG, Meyer TM, et al. (2011) Substrate specificity of protein tyrosine phosphatases 1b, rptpalpha, shp-1, and shp-2. Biochemistry 50: 2339-56.

[78] Sacco F, Perfetto L, Castagnoli L, Cesareni G (2012) The human phosphatase interactome: An intricate family portrait. FEBS Letters 586: 2732-9.

[79] Haugh J, Schneider I, Lewis J (2004) On the cross-regulation of protein tyrosine phosphatases and receptor tyrosine kinases in intracellular signaling. J Theor Biol 230: 119-132.

[80] Bollen M (2001) Combinatorial control of protein phosphatase-1. Trends in biochemical sciences 26: 426-31.

[81] Goldberg Y (1999) Protein phosphatase 2a: who shall regulate the regulator? Biochemical Pharmacology 57: 321-8.

[82] Sontag E (2001) Protein phosphatase 2a: the trojan horse of cellular signaling. Cellular Signalling 13: 7-16.

[83] Virshup DM (2000) Protein phosphatase 2a: a panoply of enzymes. Current Opinion in Cell Biology 12: 180-5.

[84] Xu Y, Xing Y, Chen Y, Chao Y, Lin Z, et al. (2006) Structure of the protein phosphatase 2a holoenzyme. Cell 127: 1239-51.

[85] Sontag E, Fedorov SA, Kamibayashi C, Robbins D, Cobb M, et al. (1993) The interaction of sv40 small tumor antigen with protein phosphatase 2aq stimulates the map kinase pathway and induces cell proliferation. Cell 75: 887-897.

[86] Hindmarsh AC, Brown PN, Grant KE, Lee SL, Serban R, et al. (2005) Sundials: Suite of nonlinear and differential/algebraic equation solvers. Acm Transactions on Mathematical Software 31: 363-396.

[87] Cambridge SB, Gnad F, Nguyen C, Bermejo JL, Kruger M, et al. (2011) Systems-wide proteomic analysis in mammalian cells reveals conserved, functional protein turnover. Journal of proteome research 10: 5275-84.

[88] Team RC (2013) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

[89] Consortium TU (2013) Update on activities at the universal protein resource (uniprot) in 2013. Nucleic Acids Research 41: D43-7.

[90] Ghaemmaghami S, Huh W, Bower K, Howson W, Belle A, et al. (2003) Global analysis of protein expression in yeast. Nature 425: 737-741.

[91] Suderman R, Deeds EJ (2013) Machines vs. ensembles: effective mapk signaling through heterogeneous sets of protein complexes. PLoS computational biology 9: e1003278.

[92] Rotin D, Margolis B, Mohammadi M, Daly RJ, Daum G, et al. (1992) Sh2 domains prevent tyrosine dephosphorylation of the egf receptor: identification of tyr992 as the high-affinity binding site for sh2 domains of phospholipase c gamma. The EMBO journal 11: 559-567.

[93] Chiang C, Harris G, Ellig C, Masters SC, Subramanian R, et al. (2001) Protein phosphatase 2a activates the proapoptotic function of bad in interleukin-3-dependent lymphoid cells by a mechanism requiring 14-3-3 dissociation. Blood 97: 1289-1297.

[94] Pederson TM, Kramer DL, Rondinone CM (2001) Serine/threonine phosphorylation of irs-1 triggers its degradation: possible regulation by tyrosine phosphorylation. Diabetes 50: 24-31.

[95] Wang Y, Nishina PM, Naggert JK (2009) Degradation of irs1 leads to impaired glucose uptake in adipose tissue of the type 2 diabetes mouse model tallyho/jng. The Journal of endocrinology 203: 65-74.

[96] Lew J, Taylor SS, Adams JA (1997) Identification of a partially rate-determining step in the catalytic mechanism of camp-dependent protein kinase: a transient kinetic study using stopped-flow fluorescence spectroscopy. Biochemistry 36: 6717-24.

[97] Grant BD, Adams JA (1996) Pre-steady-state kinetic analysis of camp-dependent protein kinase using rapid quench flow techniques. Biochemistry 35: 2022-9.

[98] Yang J, Hlavacek W (2011) Scaffold-mediated nucleation of protein signaling complexes: elementary principles. Math Biosci 232: 164-173.

[99] Oberdorf R, Kortemme T (2009) Complex topology rather than complex membership is a determinant of protein dosage sensitivity. Molecular systems biology 5: 253.

[100] Chapman SA, Asthagiri AR (2009) Quantitative effect of scaffold abundance on signal propagation. Molecular systems biology 5: 313.

[101] Levchenko A, Bruck J, Sternberg PW (2000) Scaffold proteins may biphasically affect the levels of mitogen-activated protein kinase signaling and reduce its threshold properties. Proceedings of the National Academy of Sciences of the United States of America 97: 5818-23.

[102] Bray D, Lay S (1997) Computer-based analysis of the binding steps in protein complex formation. Proceedings of the National Academy of Sciences of the United States of America 94: 13493-8.

[103] Douglass J EF, Miller CJ, Sparer G, Shapiro H, Spiegel D (2013) A comprehensive mathematical model for three-body binding equilibria. J Am Chem Soc 135: 6092-6099.

[104] Van Kanegan MJ, Strack S (2009) The protein phosphatase 2a regulatory subunits b'beta and b'delta mediate sustained trka neurotrophin receptor autophosphorylation and neuronal differentiation. Molecular and Cellular Biology 29: 662-74.

[105] Gao A, Liu B, Shi X, Jia X, Ye M, et al. (2007) Phosphatidylinositol-3 kinase/akt/p70s6k/ap-1 signaling pathway mediated benzo(a)pyrene-induced cell cycle alternation via cell cycle regulatory proteins in human embryo lung fibroblasts. Toxicology Letters 170: 30-41.

[106] Zhang W, Thompson BJ, Hietakangas V, Cohen SM (2011) Mapk/erk signaling regulates insulin sensitivity to control glucose metabolism in drosophila. PLoS Genetics 7: e1002429.

[107] Vartanian R, Masri J, Martin J, Cloninger C, Holmes B, et al. (2011) Ap-1 regulates cyclin d1 and c-myc transcription in an akt-dependent manner in response to mtor inhibition: role of aip4/itch-mediated junb degradation. Molecular cancer research : MCR 9: 115-30.

[108] Kaushansky A, Gordus A, Chang B, Rush J, MacBeath G (2008) A quantitative study of the recruitment potential of all intracellular tyrosine residues on egfr, fgfr1 and igf1r. Molecular bioSystems 4: 643-53.

[109] Cullis J, Meiri D, Sandi M, Radulovich N, Kent O, et al. (2014) The rhogef gef-h1 is required for oncogeneic ras signaling via ksr-1. Cancer Cell 25: 181-195.

[110] Padmanabhan S, Mukhopadhyay A, Naraismhan S, Tesz G, Czach M, et al. (2009) A pp2a regulatory subunit regulates c. elegans insulin/igf-i signaling by modulating akt-1 phosphorylation. Cell 136: 939-951.

[111] Sorger P (2005) A reductionist's systems biology: opinion. Curr Opin Cell Biol 17: 9-11.

[112] Chen W, Niepel M, Sorger P (2010) Classic and contemporary approaches to modeling biochemical reactions. Genes Dev 24: 1861-1875.

[113] Janes K, Lauffenburger D (2013) Models of signalling networks - what cell biologists can gain from them and give to them. J Cell Sci 126: 1913-1921.

[114] Chylek L, Harris L, Tung C, Faeder JR, Lopez C, et al. (2014) Rule-based modeling: a computational approach for studying biomolecular site dynamics in cell signaling systems. Wiley Interdiscip Rev Syst Biol Med 6: 13-36.

[115] Lois AF, Weinstein M, Ditta GS, Helinski DR (1993) Autophosphorylation and phosphatase activities of the oxygen-sensing protein fixl of rhizobium meliloti are coordinately regulated by oxygen. The Journal of Biological Chemistry 268: 4370-5.

[116] Keener J, Kustu S (1988) Protein kinase and phosphoprotein phosphatase activities of nitrogen regulatory proteins ntrb and ntrc of enteric bacteria: roles of the conserved amino-terminal domain of ntrc. Proceedings of the National Academy of Sciences of the United States of America 85: 4976-80.

[117] Mizuno T (1997) Compilation of all genes encoding two-component phosphotransfer signal transducers in the genome of escherichia coli. DNA research : an international journal for rapid publication of reports on genes and genomes 4: 161-8.

[118] Capra EJ, Perchuk BS, Lubin EA, Ashenberg O, Skerker JM, et al. (2010) Systematic dissection and trajectory-scanning mutagenesis of the molecular interface that ensures specificity of two-component signaling pathways. PLoS Genetics 6: e1001220.

[119] Capra EJ, Perchuk BS, Skerker JM, Laub MT (2012) Adaptive mutations that prevent crosstalk enable the expansion of paralogous signaling protein families. Cell 150: 222-32.

[120] Lyons SM, Prasad A (2012) Cross-talk and information transfer in mammalian and bacterial signaling. PloS one 7: e34488.

[121] Alm E, Huang K, Arkin A (2006) The evolution of two-component systems in bacteria reveals different strategies for niche adaptation. PLoS Comput Biol 2: e143.

[122] Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, et al. (2012) The pfam protein families database. Nucleic Acids Research 40: D290-301.

[123] Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal w and clustal x version 2.0. Bioinformatics 23: 2947-8.

[124] Charif D, Lobry J (2007) SeqinR 1.0-2: A contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. Structural Approaches to Sequences Evolutions: Molecules, Networks, Populations. New York: Springer.

[125] Cai SJ, Inouye M (2002) Envz-ompr interaction and osmoregulation in escherichia coli. The Journal of Biological Chemistry 277: 24155-61.

[126] Lee J, Tomchick DR, Brautigam CA, Machius M, Kort R, et al. (2008) Changes at the kina pas-a dimerization interface influence histidine kinase function. Biochemistry 47: 4051-64.

[127] Hurst LD (2002) The ka/ks ratio: diagnosing the form of sequence evolution. Trends in genetics : TIG 18: 486.

[128] Zapf J, Sen U, Madhusudan, Hoch JA, Varughese KI (2000) A transient interaction between two phosphorelay proteins trapped in a crystal lattice reveals the mechanism of molecular recognition and phosphotransfer in signal transduction. Structure 8: 851-62.

[129] Porter SL, Wadhams GH, Armitage JP (2011) Signal processing in complex chemotaxis pathways. Nature reviews Microbiology 9: 153-65.

[130] Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, et al. (2002) Network motifs: simple building blocks of complex networks. Science 298: 824-7.

[131] Shen-Orr SS, Milo R, Mangan S, Alon U (2002) Network motifs in the transcriptional regulation network of escherichia coli. Nature genetics 31: 64-8.

[132] Komiya Y, Habas R (2008) Wnt signal transduction pathways. Organogenesis 4: 68-75.

[133] Jin Z, El-Deiry WS (2005) Overview of cell death signaling pathways. Cancer biology & therapy 4: 139-63.

[134] Bhalla US, Iyengar R (1999) Emergent properties of networks of biological signaling pathways. Science 283: 381-7.

[135] Rowland MA, Deeds EJ (2014) Crosstalk and the evolution of specificity in two-component signaling. Proc Natl Acad Sci U S A 111: 5550-5555.

[136] Chang HY, Sneddon JB, Alizadeh AA, Sood R, West RB, et al. (2004) Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. PLoS biology 2: E7.

[137] Garg HG, Longaker MT, editors (2000) Scarless wound healing. New York: Dekker.

[138] Midwood KS, Williams LV, Schwarzbauer JE (2004) Tissue repair and the dynamics of the extracellular matrix. The international journal of biochemistry & cell biology 36: 1031-7.

[139] Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, et al. (2010) Towards a knowledge-based human protein atlas. Nature biotechnology 28: 1248-50.

[140] Arva NC, Talbott KE, Okoro DR, Brekman A, Qiu WG, et al. (2008) Disruption of the p53-mdm2 complex by nutlin-3 reveals different cancer cell phenotypes. Ethnicity & Disease 18: S2-1-8.

[141] Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, et al. (2012) The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 483: 603-7.

[142] Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, et al. (2012) Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature 483: 570-5.

[143] Chen JY, Lin JR, Cimprich KA, Meyer T (2012) A two-dimensional erk-akt signaling code for an ngf-triggered cell-fate decision. Molecular cell 45: 196-209.

[144] Siegal ML, Bergman A (2002) Waddington's canalization revisited: developmental stability and evolution. Proceedings of the National Academy of Sciences of the United States of America 99: 10528-32.

[145] Raman K, Wagner A (2011) Evolvability and robustness in a complex signalling circuit. Molecular bioSystems 7: 1081-92.

[146] Raman K, Wagner A (2011) The evolvability of programmable hardware. Journal of the Royal Society, Interface / the Royal Society 8: 269-81.

[147] Kirschner M, Gerhart J (1998) Evolvability. Proceedings of the National Academy of Sciences of the United States of America 95: 8420-7.

[148] Fallahi-Sichani M, Honarnejad S, Heiser LM, Gray JW, Sorger PK (2013) Metrics other than potency reveal systematic variation in responses to cancer drugs. Nature Chemical Biology 9: 708-14.

[149] Logue JS, Morrison DK (2012) Complexity in the signaling network: insights from the use of targeted inhibitors in cancer therapy. Genes & development 26: 641-50.

[150] Tovar C, Rosinski J, Filipovic Z, Higgins B, Kolinsky K, et al. (2006) Small-molecule mdm2 antagonists reveal aberrant p53 signaling in cancer: implications for therapy. Proceedings of the National Academy of Sciences of the United States of America 103: 1888-93.

[151] Huang B, Deo D, Xia M, Vassilev LT (2009) Pharmacologic p53 activation blocks cell cycle progression but fails to induce senescence in epithelial cancer cells. Molecular cancer research : MCR 7: 1497-509.

[152] Paris R, Henry RE, Stephens SJ, McBryde M, Espinosa JM (2008) Multiple p53-independent gene silencing mechanisms define the cellular response to p53 activation. Cell cycle 7: 2427-33.

[153] Satija R, Shalek AK (2014) Heterogeneity in immune responses: from populations to single cells. Trends in Immunology 35: 219-29.

[154] Feinerman O, Veiga J, Dorfman JR, Germain RN, Altan-Bonnet G (2008) Variability and robustness in t cell activation from regulated heterogeneity in protein levels. Science 321: 1081-4.

[155] Gerlach C, Rohr JC, Perie L, van Rooij N, van Heijst JW, et al. (2013) Heterogeneous differentiation patterns of individual cd8+ t cells. Science 340: 635-9.

[156] Blake WJ, Balazsi G, Kohanski MA, Isaacs FJ, Murphy KF, et al. (2006) Phenotypic consequences of promoter-mediated transcriptional noise. Molecular cell 24: 853-65.

[157] Eldar A, Elowitz MB (2010) Functional roles for noise in genetic circuits. Nature 467: 167-73.

[158] Tamborero D, Gonzalez-Perez A, Perez-Llamas C, Deu-Pons J, Kandoth C, et al. (2013) Comprehensive identification of mutational cancer driver genes across 12 tumor types. Scientific reports 3: 2650.

[159] Research W (2010) Mathematica. Champaign, IL: Wolfram Research.

[160] Qin S, Pang X, Zhou H (2011) Automated prediction of protein association rate constants. Structure 19: 1744-1751.

[161] Felsenstein J (1989) Phylip - phylogeny inference package (version 3.2). Cladistics 5: 164-166.

[162] Altschul S, Gish W, Miller W, Myers E, Lipman D (1990) Basic local alignment search tool. Journal of Molecular Biology 215: 403-410.

[163] Varughese K, Madhusudan M, Zhou X, Whiteley J, Hoch JA (1998) Formation of a novel four-helix bundle and molecular recognition sites by dimerization of a response regulator phosphotransferase. Molecular Cell 2: 485-493.

[164] Cavallo L, Kleinjung J, Fraternali F (2003) Pops: a fast algorithm for solvent accessible surface areas at atomic and residue level. Nucleic Acids Research 31: 3364-3366.

[165] Madhusudan M, Zapf J, Hoch J, Whiteley J, Xuong N, et al. (1997) A response regulatory protein with the site of phosphorylation blocked by an arginine interaction: crystal structure of spo0f from bacillus subtilis. Biochemistry 36: 12739-12745.

Appendix A

Appendix for Crosstalk and Competition in Signaling Networks

A.1 Systems of Ordinary Differential Equations

A.1.1 1-Kinase/1-Phosphatase loop with two substrates

The set of enzymatic reactions for the 1K1P loop with two substrates is as in equation 2.3 in Chapter 2:

$$S_1 + K \underset{k_{-,K,1}}{\overset{k_{+,K,1}}{\rightleftharpoons}} KS_1 \xrightarrow{k_{cat,K,1}} S_1^* + K$$

$$S_2 + K \underset{k_{-,K,2}}{\overset{k_{+,K,2}}{\rightleftharpoons}} KS_2 \xrightarrow{k_{cat,K,2}} S_2^* + K$$

$$S_1^* + P \underset{k_{-,P,1}}{\overset{k_{+,P,1}}{\rightleftharpoons}} PS_1^* \xrightarrow{k_{cat,P,1}} S_1 + P$$

$$S_2^* + P \underset{k_{-,P,2}}{\overset{k_{+,P,2}}{\rightleftharpoons}} PS_2^* \xrightarrow{k_{cat,P,2}} S_2 + P$$

Each contain three rates: rate of complex formation, (k_+), rate of complex dissociation (k_-), and catalytic rate (k_{cat}). The set of ODEs describing the free enzymes are:

$$\frac{d[K]}{dt} = -\left([S_1]\cdot[K]\cdot k_{+,K,1} + [S_2]\cdot[K]\cdot k_{+,K,2}\right) + \left([KS_1]\cdot(k_{-,K,1}+k_{cat,K,1}) + [KS_2]\cdot(k_{-,K,2}+k_{cat,K,2})\right)$$
$$\frac{d[P]}{dt} = -\left([S_1^*]\cdot[P]\cdot k_{+,P,1} + [S_2^*]\cdot[P]\cdot k_{+,P,2}\right) + \left([PS_1^*]\cdot(k_{-,P,1}+k_{cat,P,1}) + [PS_2^*]\cdot(k_{-,P,2}+k_{cat,P,2})\right)$$

The set of ODEs describing the unmodified substrates are:

$$\frac{d[S_1]}{dt} = -\left([S_1]\cdot[K]\cdot k_{+,K,1}\right) + \left([KS_1]\cdot k_{-,K,1} + [PS_1^*]\cdot k_{cat,P,1}\right)$$
$$\frac{d[S_2]}{dt} = -\left([S_2]\cdot[K]\cdot k_{+,K,2}\right) + \left([KS_2]\cdot k_{-,K,2} + [PS_2^*]\cdot k_{cat,P,2}\right)$$

The set of ODEs describing the modified substrates are:

$$\frac{d[S_1^*]}{dt} = -\left([S_1^*]\cdot[P]\cdot k_{+,P,1}\right) + \left([PS_1^*]\cdot k_{-,P,1} + [KS_1]\cdot k_{cat,K,1}\right)$$
$$\frac{d[S_2^*]}{dt} = -\left([S_2^*]\cdot[P]\cdot k_{+,P,2}\right) + \left([PS_2^*]\cdot k_{-,P,2} + [KS_2]\cdot k_{cat,K,2}\right)$$

The set of ODEs describing the enzyme-substrate complexes are:

$$\frac{d[KS_1]}{dt} = -\left([KS_1]\cdot(k_{-,K,1}+k_{cat,K,1})\right) + \left([S_1]\cdot[K]\cdot k_{+,K,1}\right)$$
$$\frac{d[KS_2]}{dt} = -\left([KS_2]\cdot(k_{-,K,2}+k_{cat,K,2})\right) + \left([S_2]\cdot[K]\cdot k_{+,K,2}\right)$$
$$\frac{d[PS_1^*]}{dt} = -\left([PS_1^*]\cdot(k_{-,P,1}+k_{cat,P,1})\right) + \left([S_1^*]\cdot[P]\cdot k_{+,P,1}\right)$$
$$\frac{d[PS_2^*]}{dt} = -\left([PS_2^*]\cdot(k_{-,P,2}+k_{cat,P,2})\right) + \left([S_2^*]\cdot[P]\cdot k_{+,P,2}\right)$$

For purposes of display in Fig. 2.3*A* of Chapter 2 we used the following values for each of the rate constants:

Parameter	Value
$k_{+,K,i}$	0.001 $\text{nM}^{-1}\cdot\text{s}^{-1}$
$k_{-,K,i}$	0.001 s^{-1}
$k_{cat,K,i}$	0.999 s^{-1}
$k_{+,P,i}$	0.001 $\text{nM}^{-1}\cdot\text{s}^{-1}$
$k_{-,P,i}$	0.001 s^{-1}
$k_{cat,P,i}$	0.999 s^{-1}

Where i = 1 or 2.

Our simulations started with the following initial concentrations:

Molecular Species	Initial Concentration
K	0 - 2 nM
P	1 nM
S_1	100 nM
S_2	0 - 20 μM

With the remaining molecular species having initial concentrations of 0. The range of initial concentrations of K and S_2 were used to vary r_1 and $[S_2]_0/K_m$, respectively, in Fig. 2.3*A* in Chapter 2.

A.1.2 1-Kinase/1-Phosphatase loop with many substrates

The set of enzymatic reactions for the 1K1P loop with many substrates is:

$$S_1 + K \underset{k_{-,K,1}}{\overset{k_{+,K,1}}{\rightleftharpoons}} KS_1 \xrightarrow{k_{cat,K,1}} S_1^* + K$$
$$S_2 + K \underset{k_{-,K,2}}{\overset{k_{+,K,2}}{\rightleftharpoons}} KS_2 \xrightarrow{k_{cat,K,2}} S_2^* + K$$
$$\vdots$$
$$S_N + K \underset{k_{-,K,N}}{\overset{k_{+,K,N}}{\rightleftharpoons}} KS_N \xrightarrow{k_{cat,K,N}} S_N^* + K$$

$$S_1^* + P \underset{k_{-,P,1}}{\overset{k_{+,P,1}}{\rightleftharpoons}} PS_1^* \xrightarrow{k_{cat,P,1}} S_1 + P$$
$$S_2^* + P \underset{k_{-,P,2}}{\overset{k_{+,P,2}}{\rightleftharpoons}} PS_2^* \xrightarrow{k_{cat,P,2}} S_2 + P$$
$$\vdots$$
$$S_N^* + P \underset{k_{-,P,N}}{\overset{k_{+,P,N}}{\rightleftharpoons}} PS_N^* \xrightarrow{k_{cat,P,N}} S_N + P$$

The set of ODEs describing the free enzymes are:

$$\begin{aligned}
\frac{d[K]}{dt} =& -([S_1]\cdot[K]\cdot k_{+,K,1} + [S_2]\cdot[K]\cdot k_{+,K,2} + \ldots + [S_N]\cdot[K]\cdot k_{+,K,N}) \\
&+ ([KS_1]\cdot(k_{-,K,1} + k_{cat,K,1}) + [KS_2]\cdot(k_{-,K,2} + k_{cat,K,2}) + \ldots + [KS_N]\cdot(k_{-,K,N} + k_{cat,K,N})) \\
\frac{d[P]}{dt} =& -([S_1^*]\cdot[P]\cdot k_{+,P,1} + [S_2^*]\cdot[P]\cdot k_{+,P,2} + \ldots + [S_N^*]\cdot[P]\cdot k_{+,P,N}) \\
&+ ([PS_1^*]\cdot(k_{-,P,1} + k_{cat,P,1}) + [PS_2^*]\cdot(k_{-,P,2} + k_{cat,P,2}) + \ldots + [PS_N^*]\cdot(k_{-,P,N} + k_{cat,P,N}))
\end{aligned}$$

The set of ODEs describing the unmodified substrates are:

$$\begin{aligned}
\frac{d[S_1]}{dt} &= -\left([S_1]\cdot[K]\cdot k_{+,K,1}\right) + \left([KS_1]\cdot k_{-,K,1} + [PS_1^*]\cdot k_{cat,P,1}\right) \\
\frac{d[S_2]}{dt} &= -\left([S_2]\cdot[K]\cdot k_{+,K,2}\right) + \left([KS_2]\cdot k_{-,K,2} + [PS_2^*]\cdot k_{cat,P,2}\right) \\
&\vdots \\
\frac{d[S_N]}{dt} &= -\left([S_N]\cdot[K]\cdot k_{+,K,N}\right) + \left([KS_N]\cdot k_{-,K,N} + [PS_N^*]\cdot k_{cat,P,N}\right)
\end{aligned}$$

The set of ODEs describing the modified substrates are:

$$\begin{aligned}
\frac{d[S_1^*]}{dt} &= -\left([S_1^*]\cdot[P]\cdot k_{+,P,1}\right) + \left([PS_1^*]\cdot k_{-,P,1} + [KS_1]\cdot k_{cat,K,1}\right) \\
\frac{d[S_2^*]}{dt} &= -\left([S_2^*]\cdot[P]\cdot k_{+,P,2}\right) + \left([PS_2^*]\cdot k_{-,P,2} + [KS_2]\cdot k_{cat,K,2}\right) \\
&\vdots \\
\frac{d[S_N^*]}{dt} &= -\left([S_N^*]\cdot[P]\cdot k_{+,P,N}\right) + \left([PS_N^*]\cdot k_{-,P,N} + [KS_N]\cdot k_{cat,K,N}\right)
\end{aligned}$$

The set of ODEs describing the enzyme-substrate complexes are:

$$\begin{aligned}
\frac{d[KS_1]}{dt} &= -\left([KS_1]\cdot(k_{-,K,1} + k_{cat,K,1})\right) + \left([S_1]\cdot[K]\cdot k_{+,K,1}\right) \\
\frac{d[KS_2]}{dt} &= -\left([KS_2]\cdot(k_{-,K,2} + k_{cat,K,2})\right) + \left([S_2]\cdot[K]\cdot k_{+,K,2}\right) \\
&\vdots \\
\frac{d[KS_N]}{dt} &= -\left([KS_N]\cdot(k_{-,K,N} + k_{cat,K,N})\right) + \left([S_N]\cdot[K]\cdot k_{+,K,N}\right) \\
\frac{d[PS_1^*]}{dt} &= -\left([PS_1^*]\cdot(k_{-,P,1} + k_{cat,P,1})\right) + \left([S_1^*]\cdot[P]\cdot k_{+,P,1}\right) \\
\frac{d[PS_2^*]}{dt} &= -\left([PS_2^*]\cdot(k_{-,P,2} + k_{cat,P,2})\right) + \left([S_2^*]\cdot[P]\cdot k_{+,P,2}\right) \\
&\vdots \\
\frac{d[PS_N^*]}{dt} &= -\left([PS_N^*]\cdot(k_{-,P,N} + k_{cat,P,N})\right) + \left([S_N^*]\cdot[P]\cdot k_{+,P,N}\right)
\end{aligned}$$

The following values for rate constants were used in the simulations presented in Fig. 2.3*B* of Chapter 2:

Parameter	Value
$k_{+,K,i}$	0.001 $\text{nM}^{-1}\cdot\text{s}^{-1}$
$k_{-,K,i}$	0.001 s^{-1}
$k_{cat,K,i}$	0.999 s^{-1}
$k_{+,P,i}$	0.001 $\text{nM}^{-1}\cdot\text{s}^{-1}$
$k_{-,P,i}$	0.001 s^{-1}
$k_{cat,P,i}$	0.999 s^{-1}

The different molecular species were initialized with concentrations:

Molecular Species	Initial Concentration
K	0 - 2 nM
P	1 nM
S_i	500 nM

$i = 1, 2, \ldots, N$

The remaining molecular species had initial concentrations of 0. The range of initial concentrations of K was used to vary the value of r_1, and N was varied to obtain the surface in Fig. 2.3*B* in Chapter 2.

A.1.3 1-Kinase/2-Phosphatase loop

The set of enzymatic reactions for the 1K2P loop is:

$$S_1 + K \underset{k_{-,K,1}}{\overset{k_{+,K,1}}{\rightleftharpoons}} KS_1 \xrightarrow{k_{cat,K,1}} S_1^* + K$$
$$S_2 + K \underset{k_{-,K,2}}{\overset{k_{+,K,2}}{\rightleftharpoons}} KS_2 \xrightarrow{k_{cat,K,2}} S_2^* + K$$
$$S_1^* + P_1 \underset{k_{-,P,1}}{\overset{k_{+,P,1}}{\rightleftharpoons}} P_1S_1^* \xrightarrow{k_{cat,P,1}} S_1 + P_1$$
$$S_2^* + P_2 \underset{k_{-,P,2}}{\overset{k_{+,P,2}}{\rightleftharpoons}} P_2S_2^* \xrightarrow{k_{cat,P,2}} S_2 + P_2$$

The set of ODEs describing the free enzymes are:

$$\frac{d[K]}{dt} = -\left([S_1]\cdot[K]\cdot k_{+,K,1} + [S_2]\cdot[K]\cdot k_{+,K,2}\right) + \left([KS_1]\cdot(k_{-,K,1}+k_{cat,K,1}) + [KS_2]\cdot(k_{-,K,2}+k_{cat,K,2})\right)$$
$$\frac{d[P_1]}{dt} = -\left([S_1^*]\cdot[P_1]\cdot k_{+,P,1}\right) + \left([P_1S_1^*]\cdot(k_{-,P,1}+k_{cat,P,1})\right)$$
$$\frac{d[P_2]}{dt} = -\left([S_2^*]\cdot[P_2]\cdot k_{+,P,2}\right) + \left([P_2S_2^*]\cdot(k_{-,P,2}+k_{cat,P,2})\right)$$

The set of ODEs describing the unmodified substrates are:

$$\frac{d[S_1]}{dt} = -\left([S_1]\cdot[K]\cdot k_{+,K,1}\right) + \left([KS_1]\cdot k_{-,K,1} + [P_1S_1^*]\cdot k_{cat,P,1}\right)$$
$$\frac{d[S_2]}{dt} = -\left([S_2]\cdot[K]\cdot k_{+,K,2}\right) + \left([KS_2]\cdot k_{-,K,2} + [P_2S_2^*]\cdot k_{cat,P,2}\right)$$

The set of ODEs describing the modified substrates are:

$$\frac{d[S_1^*]}{dt} = -\left([S_1^*]\cdot[P_1]\cdot k_{+,P,1}\right) + \left([P_1S_1^*]\cdot k_{-,P,1} + [KS_1]\cdot k_{cat,K,1}\right)$$
$$\frac{d[S_2^*]}{dt} = -\left([S_2^*]\cdot[P_2]\cdot k_{+,P,2}\right) + \left([P_2S_2^*]\cdot k_{-,P,2} + [KS_2]\cdot k_{cat,K,2}\right)$$

The set of ODEs describing the enzyme-substrate complexes are:

$$\begin{aligned}
\frac{d[KS_1]}{dt} &= -\left([KS_1]\cdot(k_{-,K,1}+k_{cat,K,1})\right)+\left([S_1]\cdot[K]\cdot k_{+,K,1}\right)\\
\frac{d[KS_2]}{dt} &= -\left([KS_2]\cdot(k_{-,K,2}+k_{cat,K,2})\right)+\left([S_2]\cdot[K]\cdot k_{+,K,2}\right)\\
\frac{d[P_1S_1^*]}{dt} &= -\left([P_1S_1^*]\cdot(k_{-,P,1}+k_{cat,P,1})\right)+\left([S_1^*]\cdot[P_1]\cdot k_{+,P,1}\right)\\
\frac{d[P_2S_2^*]}{dt} &= -\left([P_2S_2^*]\cdot(k_{-,P,2}+k_{cat,P,2})\right)+\left([S_2^*]\cdot[P_2]\cdot k_{+,P,2}\right)
\end{aligned}$$

For purposes of display in Figs. 4*A* and *B* in the main text, we used the following parameters in the model:

Parameter	Value
$k_{+,K,i}$	0.001 $\text{nM}^{-1}\cdot\text{s}^{-1}$
$k_{-,K,i}$	0.001 s^{-1}
$k_{cat,K,i}$	0.999 s^{-1}
$k_{+,P,i}$	0.001 $\text{nM}^{-1}\cdot\text{s}^{-1}$
$k_{-,P,i}$	0.001 s^{-1}
$k_{cat,P,i}$	0.999 s^{-1}

$i = 1$ or 2

Each of the molecular species in the model started with the following initial concentrations:

Molecular Species	Initial Concentration
K	1 nM
P_1	0.5-100 nM
P_2	0.5-100 nM
S_1	100 nM
S_2	0, 20 μM

The remaining molecular species had initial concentrations of 0. The range of initial concentrations for P_1 and P_2 were used to independently set r_1 and r_2, respectively, in Figs. 2.4*A* and *B* in Chapter

2. In Fig. 2.4*A* $[S_2]_0 = 0$ and in Fig. 2.4*B* $[S_2]_0 = 20$ nM.

A.1.4 2-Kinase/1-Phosphatase loop

The set of enzymatic reactions for the 2K1P loop is:

$$
\begin{aligned}
S_1 + K_1 &\underset{k_{-,K,1}}{\overset{k_{+,K,1}}{\rightleftharpoons}} K_1S_1 \xrightarrow{k_{cat,K,1}} S_1^* + K_1 \\
S_2 + K_2 &\underset{k_{-,K,2}}{\overset{k_{+,K,2}}{\rightleftharpoons}} K_2S_2 \xrightarrow{k_{cat,K,2}} S_2^* + K_2 \\
S_1^* + P &\underset{k_{-,P,1}}{\overset{k_{+,P,1}}{\rightleftharpoons}} PS_1^* \xrightarrow{k_{cat,P,1}} S_1 + P \\
S_2^* + P &\underset{k_{-,P,2}}{\overset{k_{+,P,2}}{\rightleftharpoons}} PS_2^* \xrightarrow{k_{cat,P,2}} S_2 + P
\end{aligned}
$$

The set of ODEs describing the free enzymes are:

$$
\begin{aligned}
\frac{d[K_1]}{dt} &= -\left([S_1]\cdot[K_1]\cdot k_{+,K,1}\right) + \left([K_1S_1]\cdot(k_{-,K,1} + k_{cat,K,1})\right) \\
\frac{d[K_2]}{dt} &= -\left([S_2]\cdot[K_2]\cdot k_{+,K,2}\right) + \left([K_2S_2]\cdot(k_{-,K,2} + k_{cat,K,2})\right) \\
\frac{d[P]}{dt} &= -\left([S_1^*]\cdot[P]\cdot k_{+,P,1} + [S_2^*]\cdot[P]\cdot k_{+,P,2}\right) + \left([PS_1^*]\cdot(k_{-,P,1} + k_{cat,P,1}) + [PS_2^*]\cdot(k_{-,P,2} + k_{cat,P,2})\right)
\end{aligned}
$$

The set of ODEs describing the unmodified substrates are:

$$
\begin{aligned}
\frac{d[S_1]}{dt} &= -\left([S_1]\cdot[K_1]\cdot k_{+,K,1}\right) + \left([K_1S_1]\cdot k_{-,K,1} + [PS_1^*]\cdot k_{cat,P,1}\right) \\
\frac{d[S_2]}{dt} &= -\left([S_2]\cdot[K_2]\cdot k_{+,K,2}\right) + \left([K_2S_2]\cdot k_{-,K,2} + [PS_2^*]\cdot k_{cat,P,2}\right)
\end{aligned}
$$

The set of ODEs describing the modified substrates are:

$$
\begin{aligned}
\frac{d[S_1^*]}{dt} &= -\left([S_1^*]\cdot[P]\cdot k_{+,P,1}\right) + \left([PS_1^*]\cdot k_{-,P,1} + [K_1S_1]\cdot k_{cat,K,1}\right) \\
\frac{d[S_2^*]}{dt} &= -\left([S_2^*]\cdot[P]\cdot k_{+,P,2}\right) + \left([PS_2^*]\cdot k_{-,P,2} + [K_2S_2]\cdot k_{cat,K,2}\right)
\end{aligned}
$$

The set of ODEs describing the enzyme-substrate complexes are:

$$\frac{d[K_1S_1]}{dt} = -\left([K_1S_1]\cdot(k_{-,K,1}+k_{cat,K,1})\right)+\left([S_1]\cdot[K_1]\cdot k_{+,K,1}\right)$$
$$\frac{d[K_2S_2]}{dt} = -\left([K_2S_2]\cdot(k_{-,K,2}+k_{cat,K,2})\right)+\left([S_2]\cdot[K_2]\cdot k_{+,K,2}\right)$$
$$\frac{d[PS_1^*]}{dt} = -\left([PS_1^*]\cdot(k_{-,P,1}+k_{cat,P,1})\right)+\left([S_1^*]\cdot[P]\cdot k_{+,P,1}\right)$$
$$\frac{d[PS_2^*]}{dt} = -\left([PS_2^*]\cdot(k_{-,P,2}+k_{cat,P,2})\right)+\left([S_2^*]\cdot[P]\cdot k_{+,P,2}\right)$$

For purposes of display in Figs. 2.4*A* and *C* in Chapter 2 we used the following parameters:

Parameter	Value
$k_{+,K,i}$	0.001 $\text{nM}^{-1}\cdot\text{s}^{-1}$
$k_{-,K,i}$	0.001 s^{-1}
$k_{cat,K,i}$	0.999 s^{-1}
$k_{+,P,i}$	0.001 $\text{nM}^{-1}\cdot\text{s}^{-1}$
$k_{-,P,i}$	0.001 s^{-1}
$k_{cat,P,i}$	0.999 s^{-1}

$i = 1$ or 2

Each molecular species were initialized at the following concentrations:

Molecular Species	Initial Concentration
K_1	0 - 2 nM
K_2	0 - 2 nM
P	1 nM
S_1	100 nM
S_2	0, 20 μM

The remaining molecular species had initial concentrations of 0. The range of initial concentrations of K_1 and K_2 were used to set the values of r_1 and r_2, respectively, in Figs. 2.4*A* and *C* in Chapter 2. In Fig. 2.4*A*, $[S_2]_0 = 0$ and in Fig. 2.4*C*, $[S_2]_0 = 20$ nM.

A.1.5 Cascade with multiple phosphatases

The set of kinase enzymatic reactions for the cascade with multiple phosphatases is:

$$
\begin{aligned}
S_1 + K &\underset{k_{-,K,1}}{\overset{k_{+,K,1}}{\rightleftharpoons}} KS_1 \xrightarrow{k_{cat,K,1}} S_1^* + K \\
S_2 + S_1^* &\underset{k_{-,K,2}}{\overset{k_{+,K,2}}{\rightleftharpoons}} S_1^* S_2 \xrightarrow{k_{cat,K,2}} S_2^* + S_1^* \\
&\vdots \\
S_N + S_{N-1}^* &\underset{k_{-,K,N}}{\overset{k_{+,K,N}}{\rightleftharpoons}} S_{N-1}^* S_N \xrightarrow{k_{cat,K,N}} S_N^* + S_{N-1}^*
\end{aligned}
$$

Note that K is the input kinase and S_i^* serves as the kinase for S_{i+1}. The set of phosphatase enzymatic reactions is:

$$\begin{aligned}
S_1^* + P_1 &\underset{k_{-,P,1}}{\overset{k_{+,P,1}}{\rightleftharpoons}} P_1S_1^* \xrightarrow{k_{cat,P,1}} S_1 + P_1 \\
S_2^* + P_2 &\underset{k_{-,P,2}}{\overset{k_{+,P,2}}{\rightleftharpoons}} P_2S_2^* \xrightarrow{k_{cat,P,2}} S_2 + P_2 \\
&\vdots \\
S_N^* + P_N &\underset{k_{-,P,N}}{\overset{k_{+,P,N}}{\rightleftharpoons}} P_NS_N^* \xrightarrow{k_{cat,P,N}} S_N + P_N
\end{aligned}$$

The set of ODEs describing the free enzymes are:

$$\begin{aligned}
\frac{d[K]}{dt} &= -\left([S_1]\cdot[K]\cdot k_{+,K,1}\right) + \left([KS_1]\cdot(k_{-,K,1}+k_{cat,K,1})\right) \\
\frac{d[P_1]}{dt} &= -\left([S_1^*]\cdot[P_1]\cdot k_{+,P,1}\right) + \left([P_1S_1^*]\cdot(k_{-,P,1}+k_{cat,P,1})\right) \\
\frac{d[P_i]}{dt} &= -\left([S_i^*]\cdot[P_i]\cdot k_{+,P,i}\right) + \left([P_iS_i^*]\cdot(k_{-,P,i}+k_{cat,P,i})\right) \\
\frac{d[P_N]}{dt} &= -\left([S_N^*]\cdot[P_N]\cdot k_{+,P,N}\right) + \left([P_NS_N^*]\cdot(k_{-,P,N}+k_{cat,P,N})\right)
\end{aligned}$$

The set of ODEs describing the unmodified substrates are:

$$\begin{aligned}
\frac{d[S_1]}{dt} &= -\left([S_1]\cdot[K]\cdot k_{+,K,1}\right) + \left([KS_1]\cdot k_{-,K,1} + [P_1S_1^*]\cdot k_{cat,P,1}\right) \\
\frac{d[S_i]}{dt} &= -\left([S_i]\cdot[S_{i-1}^*]\cdot k_{+,K,i}\right) + \left([S_{i-1}^*S_i]\cdot k_{-,K,i} + [P_iS_i^*]\cdot k_{cat,P,i}\right) \\
\frac{d[S_N]}{dt} &= -\left([S_N]\cdot[S_{N-1}^*]\cdot k_{+,K,N}\right) + \left([S_{N-1}^*S_N]\cdot k_{-,K,N} + [P_NS_N^*]\cdot k_{cat,P,N}\right)
\end{aligned}$$

The set of ODEs describing the modified substrates are:

$$\begin{aligned}
\frac{d[S_1^*]}{dt} =& -\left([S_1^*]\cdot[P]\cdot k_{+,K,1}\right) + [S_2]\cdot[S_1^*]\cdot k_{+,K,2}) \\
&+ \left([P_1S_1^*]\cdot k_{-,P,1} + [KS_1]\cdot k_{cat,K,1} + [S_1^*S_2]\cdot(k_{-,K,2}+k_{cat,K,2})\right) \\
\frac{d[S_i^*]}{dt} =& -\left([S_i^*]\cdot[P_i]\cdot k_{+,K,i}\right) + [S_{i+1}]\cdot[S_i^*]*k_{+,K,i+1}) \\
&+ \left([P_iS_i^*]\cdot k_{-,P,i} + [S_{i-1}^*S_i]\cdot k_{cat,K,i} + [S_i^*S_{i+1}]\cdot(k_{-,K,i+1}+k_{cat,K,i+1})\right) \\
\frac{d[S_N^*]}{dt} =& -\left([S_N^*]\cdot[P_N]\cdot k_{+,K,N}\right) + \left([P_NS_N^*]\cdot k_{-,P,N} + [S_{N-1}^*S_N]\cdot k_{cat,K,N}\right)
\end{aligned}$$

The set of ODEs describing the enzyme-substrate complexes are:

$$\begin{aligned}
\frac{d[KS_1]}{dt} &= -\left([KS_1]\cdot(k_{-,K,1}+k_{cat,K,1})\right) + \left([S_1]\cdot[K]\cdot k_{+,K,1}\right) \\
\frac{d[S_{i-1}^*S_i]}{dt} &= -\left([S_{i-1}^*S_i]\cdot(k_{-,K,i}+k_{cat,K,i})\right) + \left([S_i]\cdot[S_{i-1}^*]\cdot k_{+,K,i}\right) \\
\frac{d[S_{N-1}^*S_N]}{dt} &= -\left([S_{N-1}^*S_N]\cdot(k_{-,K,N}+k_{cat,K,N})\right) + \left([S_N]\cdot[S_{N-1}^*]\cdot k_{+,K,N}\right) \\
\frac{d[P_1S_1^*]}{dt} &= -\left([P_1S_1^*]\cdot(k_{-,P,1}+k_{cat,P,1})\right) + \left([S_1^*]\cdot[P_1]\cdot k_{+,P,1}\right) \\
\frac{d[P_iS_i^*]}{dt} &= -\left([P_iS_i^*]\cdot(k_{-,P,i}+k_{cat,P,i})\right) + \left([S_i^*]\cdot[P_i]\cdot k_{+,P,i}\right) \\
\frac{d[P_NS_N^*]}{dt} &= -\left([P_NS_N^*]\cdot(k_{-,P,N}+k_{cat,P,N})\right) + \left([S_N^*]\cdot[P_N]\cdot k_{+,P,N}\right)
\end{aligned}$$

Where $i = 2,\ldots,N-1$.

For purposes of display in Fig. 2.5 in the main text, we used the following parameters:

Parameter	Value
$k_{+,K,i}$	$0.001\ \text{nM}^{-1}\cdot\text{s}^{-1}$
$k_{-,K,i}$	$10^{i-8}\ \text{s}^{-1}$
$k_{cat,K,i}$	$0.999\cdot10^{i-5}\ \text{s}^{-1}$
$k_{+,P,i}$	$0.001\ \text{nM}^{-1}\cdot\text{s}^{-1}$
$k_{-,P,i}$	$10^{i-8}\ \text{s}^{-1}$
$k_{cat,P,i}$	$0.999\cdot10^{i-5}\ \text{s}^{-1}$

$$i = 1, 2, \ldots, N$$

The k_{cat}'s and k_{-}'s were calculated as $0.999 \cdot 10^{i-5}$ s^{-1} and 10^{i-8} s^{-1}, respectively; the kinetic parameters of reaction i in the cascade were thus varied so that each substrate concentration was $10 \cdot K_m$ in respect to its kinase and phosphatase.

The molecular species in the system started with the following initial concentrations:

Molecular Species	Initial Concentration
K	10^{-18} - 0.1 nM
P_i	0.01 nM
S_1	1 nM
S_2	10 nM
$\vdots$	$\vdots$
S_i	$10 \cdot S_{i-1}$
$\vdots$	$\vdots$
S_N	$10 \cdot S_{N-1}$

$$i = 1, 2, \ldots, N$$

The remaining molecular species had initial concentrations of 0. We systematically increased the initial concentration of the S_i's ($[S_i]_0 = 10 \cdot [S_{i-1}]_0$); since S^*_{i-1} is the kinase for S_i, this ensured that all substrates were at higher concentrations than their enzymes. The range of initial concentrations of K were used to vary the value of r in Fig. 2.5.

A.1.6 Cascade with a single phosphatase

The set of kinase enzymatic reactions for the cascade with a single phosphatase is:

$$
\begin{aligned}
S_1 + K \quad &\underset{k_{-,K,1}}{\overset{k_{+,K,1}}{\rightleftharpoons}} \quad KS_1 \quad \xrightarrow{k_{cat,K,1}} \quad S_1^* + K \\
S_2 + S_1^* \quad &\underset{k_{-,K,2}}{\overset{k_{+,K,2}}{\rightleftharpoons}} \quad S_1^* S_2 \quad \xrightarrow{k_{cat,K,2}} \quad S_2^* + S_1^* \\
&\vdots \\
S_N + S_{N-1}^* \quad &\underset{k_{-,K,N}}{\overset{k_{+,K,N}}{\rightleftharpoons}} \quad S_{N-1}^* S_N \quad \xrightarrow{k_{cat,K,N}} \quad S_N^* + S_{N-1}^*
\end{aligned}
$$

The set of phosphatase enzymatic reactions is:

$$
\begin{aligned}
S_1^* + P \quad &\underset{k_{-,P,1}}{\overset{k_{+,P,1}}{\rightleftharpoons}} \quad PS_1^* \quad \xrightarrow{k_{cat,P,1}} \quad S_1 + P \\
S_2^* + P \quad &\underset{k_{-,P,2}}{\overset{k_{+,P,2}}{\rightleftharpoons}} \quad PS_2^* \quad \xrightarrow{k_{cat,P,2}} \quad S_2 + P \\
&\vdots \\
S_N^* + P \quad &\underset{k_{-,P,N}}{\overset{k_{+,P,N}}{\rightleftharpoons}} \quad PS_N^* \quad \xrightarrow{k_{cat,P,N}} \quad S_N + P
\end{aligned}
$$

The set of ODEs describing free enzymes are:

$$
\begin{aligned}
\frac{d[K]}{dt} &= -\left([S_1]\cdot[K]\cdot k_{+,K,1}\right) + \left([KS_1]\cdot(k_{-,K,1}+k_{cat,K,1})\right) \\
\frac{d[P]}{dt} &= -\left([S_1^*]\cdot[P]\cdot k_{+,P,1} + [S_2^*]\cdot k_{+,P,2} + \ldots + [S_N^*]\cdot k_{+,P,N}\right) \\
&\quad + \left([PS_1^*]\cdot(k_{-,P,1}+k_{cat,P,1}) + [PS_2^*]\cdot(k_{-,P,2}+k_{cat,P,2}) + \ldots + [PS_N^*]\cdot(k_{-,P,N}+k_{cat,P,N})\right)
\end{aligned}
$$

The set of ODEs describing unmodified substrates are:

$$\begin{aligned}
\frac{d[S_1]}{dt} &= -\left([S_1]\cdot[K]\cdot k_{+,K,1}\right) + \left([KS_1]\cdot k_{-,K,1} + [PS_1^*]\cdot k_{cat,P,1}\right) \\
\frac{d[S_i]}{dt} &= -\left([S_i]\cdot[S_{i-1}^*]\cdot k_{+,K,i}\right) + \left([S_{i-1}^* S_i]\cdot k_{-,K,i} + [PS_i^*]\cdot k_{cat,P,i}\right) \\
\frac{d[S_N]}{dt} &= -\left([S_N]\cdot[S_{N-1}^*]\cdot k_{+,K,N}\right) + \left([S_{N-1}^* S_N]\cdot k_{-,K,N} + [PS_N^*]\cdot k_{cat,P,N}\right)
\end{aligned}$$

The set of ODEs describing modified substrates are:

$$\begin{aligned}
\frac{d[S_1^*]}{dt} &= -\left([S_1^*]\cdot[P]\cdot k_{+,K,1}\right) + [S_2]\cdot[S_1^*] * k_{+,K,2} \\
&\quad + \left([PS_1^*]\cdot k_{-,P,1} + [KS_1]\cdot k_{cat,K,1} + [S_1^* S_2]\cdot(k_{-,K,2} + k_{cat,K,2})\right) \\
\frac{d[S_i^*]}{dt} &= -\left([S_i^*]\cdot[P]\cdot k_{+,K,i}\right) + [S_{i+1}]\cdot[S_i^*] * k_{+,K,i+1} \\
&\quad + \left([PS_i^*]\cdot k_{-,P,i} + [S_{i-1}^* S_i]\cdot k_{cat,K,i} + [S_i^* S_{i+1}]\cdot(k_{-,K,i+1} + k_{cat,K,i+1})\right) \\
\frac{d[S_N^*]}{dt} &= -\left([S_N^*]\cdot[P]\cdot k_{+,K,N}\right) + \left([PS_N^*]\cdot k_{-,P,N} + [S_{N-1}^* S_N]\cdot k_{cat,K,N}\right)
\end{aligned}$$

The set of ODEs describing enzyme-substrate complexes are:

$$\begin{aligned}
\frac{d[KS_1]}{dt} &= -\left([KS_1]\cdot(k_{-,K,1} + k_{cat,K,1})\right) + \left([S_1]\cdot[K]\cdot k_{+,K,1}\right) \\
\frac{d[PS_1^*]}{dt} &= -\left([PS_1^*]\cdot(k_{-,P,1} + k_{cat,P,1})\right) + \left([S_1^*]\cdot[P]\cdot k_{+,P,1}\right) \\
\frac{d[S_{i-1}^* S_i]}{dt} &= -\left([S_{i-1}^* S_i]\cdot(k_{-,K,i} + k_{cat,K,i})\right) + \left([S_i]\cdot[S_{i-1}^*]\cdot k_{+,K,i}\right) \\
\frac{d[PS_i^*]}{dt} &= -\left([PS_i^*]\cdot(k_{-,P,i} + k_{cat,P,i})\right) + \left([S_i^*]\cdot[P]\cdot k_{+,P,i}\right) \\
\frac{d[S_{N-1}^* S_N]}{dt} &= -\left([S_{N-1}^* S_N]\cdot(k_{-,K,N} + k_{cat,K,N})\right) + \left([S_N]\cdot[S_{N-1}^*]\cdot k_{+,K,N}\right) \\
\frac{d[P_N S_N^*]}{dt} &= -\left([PS_N^*]\cdot(k_{-,P,N} + k_{cat,P,N})\right) + \left([S_N^*]\cdot[P]\cdot k_{+,P,N}\right)
\end{aligned}$$

Where $i = 2,\ldots,N-1$.

For purposes of display in Fig. 2.5, we used the following parameters:

Parameter	Value
$k_{+,K,i}$	0.001 $\text{nM}^{-1}\cdot\text{s}^{-1}$
$k_{-,K,i}$	(10^{i-8}) s^{-1}
$k_{cat,K,i}$	$(0.999 \cdot 10^{i-5})$ s^{-1}
$k_{+,P,i}$	0.001 $\text{nM}^{-1}\cdot\text{s}^{-1}$
$k_{-,P,i}$	(10^{i-8}) s^{-1}
$k_{cat,P,i}$	$(0.999 \cdot 10^{i-5})$ s^{-1}

The k_{cat}'s and k_{-}'s were calculated as in section A.1.5.

The molecular species were initialized at the following concentrations:

Molecular Species	Initial Concentration
K	10^{-18} - 0.1 nM
P	0.01 nM
S_1	1 nM
S_2	10 nM
$\vdots$	$\vdots$
S_i	$10 \cdot S_{i-1}$
$\vdots$	$\vdots$
S_N	$10 \cdot S_\text{N-1}$

$i = 1, 2, \ldots, N$

Remaining molecular species were set with initial concentrations of 0. Increasing the initial concentrations of S_i ensured that $[S_i]_0 = 10 \cdot [S_{i-1}]_0$ since S^*_{i-1} is the kinase for S_i to ensure that the concentration of substrates were larger than the concentrations of their respective kinases. The range of initial concentrations of K were used to vary the value of r in Fig. 2.5.

A.2 Analytical results for the 1-Kinase/1-Phosphatase loop

A.2.1 Mutual inhibition for competitive substrates

Here we will show that the 1K1P loop displays behavior dependent on r without regard for other parameters. The enzymatic reactions for an enzyme with two substrates can be written as:

$$E + S_1 \underset{k_{-,E,1}}{\overset{k_{+,E,1}}{\rightleftharpoons}} ES_1 \xrightarrow{k_{cat,E,1}} E + S_1^*$$
$$E + S_2 \underset{k_{-,E,2}}{\overset{k_{+,E,2}}{\rightleftharpoons}} ES_2 \xrightarrow{k_{cat,K,2}} E + S_2^*$$

with $E = K$ or P. The Michaelis-Menten constant and maximum velocity of the enzyme for either substrate are defined as:

$$K_{m,E,x} \equiv \frac{k_{-,E,x} + k_{cat,E,x}}{k_{+,E,x}}$$
$$V_{max,E,x} \equiv k_{cat,E,x}[E]_0$$

We can obtain the following kinetic equations:

$$\frac{d[ES_1]}{dt} = [E][S_1]k_{+,E,1} - [ES_1](k_{-,E,1} + k_{cat,E,1}) \quad \text{(A.2.1.1)}$$
$$\frac{d[ES_2]}{dt} = [E][S_2]k_{+,E,2} - [ES_2](k_{-,E,2} + k_{cat,E,2}) \quad \text{(A.2.1.2)}$$
$$\frac{d[S_1^*]}{dt} = k_{cat,E,1}[ES_1] \quad \text{(A.2.1.3)}$$

We also have the conservation of mass:

$$[E]_0 = [E] + [ES_1] + [ES_2] \quad \text{(A.2.1.4)}$$

Assuming pseudo–steady state for the enzymatic reactions, from equations A.2.1.1 and A.2.1.2 we get:

$$[ES_1] = \frac{[E][S_1]}{K_{m,E,1}}$$
$$[ES_2] = \frac{[E][S_2]}{K_{m,E,2}}$$

both of which can be substituted into equation A.2.1.4:

$$[E]_0 = [E]\left(1 + \frac{[S_1]}{K_{m,E,1}} + \frac{[S_2]}{K_{m,E,2}}\right)$$
$$[E] = \frac{[E]_0}{1 + \frac{[S_1]}{K_{m,E,1}} + \frac{[S_2]}{K_{mE,,2}}}$$

$$[ES_1] = \frac{[E]_0[S_1]}{[S_1] + K_{m,E,1}\left(1 + \frac{[S_2]}{K_{m,E,2}}\right)}$$

This can be substituted into A.2.1.3 to arrive at:

$$\frac{d[S_1^*]}{dt} = \frac{V_{max,E,1}[S_1]}{\alpha_{E,1}K_{m,E,1} + [S_1]} \tag{A.2.1.5}$$
$$\alpha_{E,1} \equiv 1 + \frac{[S_2]}{K_{m,E,2}}$$

where $\alpha_{E,1}$ is the inhibitory constant for S_2 competition with S_1 for E.

A.2.2 Steady-state solution for $[S_1^*]$

As Goldbeter and Koshland originally noted, for a futile cycle at steady state we will have $d[S_1^*]/dt = d[S_1]/dt$ [26]. Given A.2.1.5, for the 1K1P loop with two substrates this yields:

$$\frac{V_{max,K,1}[S_1]}{\alpha_{K,1}K_{m,K,1} + [S_1]} = \frac{V_{max,P,1}[S_1^*]}{\alpha_{P,1}K_{m,P,1} + [S_1^*]} \tag{A.2.2.1}$$

Following the standard Michaelis-Menten assumptions [26, 57], we have that $[S_i]_0 >> [K]_0, [P]_0$. This gives us $[S_1]_0 = [S_1] + [S_1^*]$, which can be substituted into A.2.2.1:

$$\frac{V_{max,K,1}([S_1]_0 - [S_1^*])}{\alpha_{K,1}K_{m,K,1} + ([S_1]_0 - [S_1^*])} = \frac{V_{max,P,1}[S_1^*]}{\alpha_{P,1}K_{m,P,1} + [S_1^*]}$$

Dividing both sides by $[S_1]_0$, we get:

$$\frac{V_{max,K,1}(1 - S_1^*)}{\alpha_{K,1}K_{K,1} + (1 - S_1^*)} = \frac{V_{max,P,1}S_1^*}{\alpha_{P,1}K_{P,1} + S_1^*} \tag{A.2.2.2}$$

$$K_{K,1} \equiv \frac{K_{m,K,1}}{[S_1]_0}, K_{P,1} \equiv \frac{K_{m,P,1}}{[S_1]_0}$$
$$S_1 \equiv \frac{[S_1]}{[S_1]_0}, S_1^* \equiv \frac{[S_1^*]}{[S_1]_0}$$

We can expand A.2.2.2:

$$\alpha_{P,1}V_{max,K,1}K_{P,1} - \alpha_{P,1}V_{max,K,1}K_{P,1}S_1^* + V_{max,K,1}S_1^* - V_{max,K,1}(S_1^*)^2$$
$$= \alpha_{K,1}V_{max,P,1}K_{K,1}S_1^* + V_{max,P,1}S_1^* - V_{max,P,1}(S_1^*)^2$$

Dividing both sides by $V_{max,P,1}$, we get:

$$r_1\alpha_{P,1}K_{P,1} - r_1\alpha_{P,1}K_{P,1}S_1^* + r_1S_1^* - r_1(S_1^*)^2 = \alpha_{K,1}K_{K,1}S_1^* + S_1^* - (S_1^*)^2$$

$$r_1 \equiv \frac{V_{max,K,1}}{V_{max,P,1}}$$

which can be simplified to:

$$(1 - r_1)(S_1^*)^2 + ((r_1 - 1) - (\alpha_{K,1}K_{K,1} + r_1\alpha_{P,1}K_{P,1}))S_1^* + r_1\alpha_{P,1}K_{P,1} = 0 \tag{A.2.2.3}$$

Solving for S_1^*:

$$S_1^* = \frac{(r_1-1)-(\alpha_{K,1}K_{K,1}+r_1\alpha_{P,1}K_{P,1})+\sqrt{((r_1-1)-(\alpha_{K,1}K_{K,1}+r_1\alpha_{P,1}K_{P,1}))^2+4(r_1-1)r_1\alpha_{P,1}K_{P,1}}}{2(r_1-1)} \tag{A.2.2.4}$$

There are two important things to note about this solution. For one, the above equation is valid for $r_1 > 0$; at $r_1 = 0$ one needs to take the other branch of the solution (i.e. the branch in which the square root term is subtracted in the numerator). Also, at $r_1 = 1$, A.2.2.4 has a nonessential singularity. To obtain the behavior at $r_1 = 1$, we see A.2.2.3 becomes:

$$-(\alpha_{K,1}K_{K,1}+\alpha_{P,1}K_{P,1})S_1^* + \alpha_{P,1}K_{P,1} = 0 \tag{A.2.2.5}$$

giving us S_1^* for $r_1 = 1$:

$$S_1^* = \frac{\alpha_{P,1}K_{P,1}}{\alpha_{K,1}K_{K,1}+\alpha_{P,1}K_{P,1}} \tag{A.2.2.6}$$

A.2.3 dS_1^*/dS_2^* is always positive

We wish to show that $\frac{dS_1^*}{dS_2^*} > 0$ regardless of the values of any parameter. This would indicate that the ultrasensitivity of S_2 transfers to S_1 (*i.e.*, since S_2^* will decrease as $[S_2]_0$ increases for $r_2 < 1$, S_1^* would also decrease). To do so we notice that, by the chain rule:

$$\frac{dS_1^*}{dS_2^*} = \frac{\partial S_1^*}{\partial \alpha_{K,1}} \cdot \frac{d\alpha_{K,1}}{dS_2^*} + \frac{\partial S_1^*}{\partial \alpha_{P,1}} \cdot \frac{d\alpha_{P,1}}{dS_2^*} \tag{A.2.3.1}$$

This is because S_1^* is a function of $\alpha_{K,1}$, $\alpha_{P,1}$, r_1, and a vector of positive constants A.2.2.4. Each of the α terms are, in turn, functions of S_2^*.

We will explore the signs of each component of A.2.3.1 to show that $\frac{dS_1^*}{dS_2^*} > 0$. Using Mathematica

[159], we can obtain the partial derivative of A.2.2.4 with respect to $\alpha_{K,1}$ at $r_1 \neq 1$:

$$\frac{\partial S_1^*}{\partial \alpha_{K,1}} = \frac{-K_{K,1} + \frac{K_{K,1}x}{\sqrt{x^2+y}}}{2(r_1-1)} \tag{A.2.3.2}$$

Where:

$$x \equiv -((r_1-1) - (\alpha_{K,1}K_{K,1} + r_1\alpha_{P,1}K_{P,1})), \qquad y \equiv 4(r_1-1)r_1\alpha_{P,1}K_{P,1} \tag{A.2.3.3}$$

Factoring out $\frac{-K_{K,1}}{\sqrt{x^2+y}}$ we obtain:

$$\frac{\partial S_1^*}{\partial \alpha_{K,1}} = \frac{-K_{K,1}}{\sqrt{x^2+y}} \cdot \frac{-x+\sqrt{x^2+y}}{2(r_1-1)} \tag{A.2.3.4}$$

Notice that the second term in A.2.3.4 is the expression for S_1^*, simplifying $\frac{dS_1^*}{d\alpha_{K,1}}$ to:

$$\frac{\partial S_1^*}{\partial \alpha_{K,1}} = \frac{-K_{K,1}}{\sqrt{x^2+y}} S_1^* \tag{A.2.3.5}$$

Note that $K_{K,1}$, S_1^* and $\sqrt{x^2+y}$ are all positive, making $\frac{\partial S_1^*}{\partial \alpha_{K,1}} < 0$ for $r_1 \neq 1$. We can also demonstrate this for $r_1 = 1$ by taking the partial derivative of A.2.2.6 with respect to $\alpha_{K,1}$:

$$\frac{\partial S_1^*}{\partial \alpha_{K,1}} = \frac{-\alpha_{P,1}K_{K,1}K_{P,1}}{(\alpha_{K,1}K_{K,1} + \alpha_{P,1}K_{P,1})^2} \tag{A.2.3.6}$$

Which is clearly negative, demonstrating that $\frac{\partial S_1^*}{\partial \alpha_{K,1}} < 0$ for any set of parameters.

Next it can be shown that $\frac{\partial S_1^*}{\partial \alpha_{P,1}} > 0$. We can obtain an expression the partial derivative of A.2.2.4 with respect to $\alpha_{P,1}$ at $r_1 \neq 1$ with Mathematica [159] :

$$\frac{\partial S_1^*}{\partial \alpha_{P,1}} = \frac{-r_1K_{P,1} + \frac{2(r_1-1)r_1K_{P,1} + r_1K_{P,1}x}{\sqrt{x^2+y}}}{2(r_1-1)} \tag{A.2.3.7}$$

By factoring out $\frac{r_1 K_{P,1}}{\sqrt{x^2+y}}$ we get:

$$\frac{\partial S_1^*}{\partial \alpha_{P,1}} = \frac{r_1 K_{P,1}}{\sqrt{x^2+y}} \left(\frac{2(r_1-1)+x-\sqrt{x^2+y}}{2(r_1-1)} \right) \tag{A.2.3.8}$$

Notice that the second term in A.2.3.8 is the expression for $1 - S_1^*$, simplifying $\frac{\partial S_1^*}{\partial \alpha_{P,1}}$ to:

$$\frac{\partial S_1^*}{\partial \alpha_{P,1}} = \frac{r_1 K_{P,1}}{\sqrt{x^2+y}} (1 - S_1^*) \tag{A.2.3.9}$$

We can easily see that A.2.3.9 is positive, confirming $\frac{\partial S_1^*}{\partial \alpha_{P,1}} > 0$ for $r_1 \neq 1$. We can also demonstrate this for $r_1 = 1$ by taking the partial derivative of A.2.2.6 with respect to $\alpha_{P,1}$:

$$\frac{\partial S_1^*}{\partial \alpha_{P,1}} = \frac{\alpha_{K,1} K_{K,1} K_{P,1}}{(\alpha_{K,1} K_{K,1} + \alpha_{P,1} K_{P,1})^2} \tag{A.2.3.10}$$

Which is clearly positive, demonstrating that $\frac{\partial S_1^*}{\partial \alpha_{P,1}} > 0$ for any set of parameters.

It is easy to show that $\frac{d\alpha_{K,1}}{dS_2^*} < 0$:

$$\begin{aligned} \alpha_{K,1} &= 1 + \frac{[S_2]}{K_{m,K,2}} \\ &= 1 + \frac{[S_2]_0 - [S_2^*]}{K_{m,K,2}} \\ &= 1 + \frac{1 - S_2^*}{K_{K,2}} \\ \frac{d\alpha_{K,1}}{dS_2^*} &= -\frac{1}{K_{K,2}} < 0 \end{aligned}$$

Similarly, we can show $\frac{d\alpha_{P,1}}{dS_2^*} > 0$:

$$\begin{aligned}\alpha_{P,1} &= 1 + \frac{[S_2^*]}{K_{m,P,2}} \\ &= 1 + \frac{S_2^*}{K_{P,2}} \\ \frac{d\alpha_{P,1}}{dS_2^*} &= \frac{1}{K_{P,2}} > 0\end{aligned}$$

Now we have determined the behaviors of each component of the two implementations of the chain rules presented in A.2.3.1 for all values of r_1 and r_2. When we refer back to the chain rule (A.2.3.1) we notice that both terms are positive:

$$\begin{aligned}\frac{dS_1^*}{dS_2^*} &= \frac{\partial S_1^*}{\partial \alpha_{K,1}} \cdot \frac{d\alpha_{K,1}}{dS_2^*} + \frac{\partial S_1^*}{\partial \alpha_{P,1}} \cdot \frac{d\alpha_{P,1}}{dS_2^*} \\ \frac{dS_1^*}{dS_2^*} &= (-)(-) + (+)(+)\end{aligned}$$

This means that changes in S_1^* upon increases in S_2^* will always be positive. The increase in ultra-sensitivity of S_2^* is thus transferred to S_1 regardless of the values of the other parameters.

A.3 Analytical results for the 1-Kinase/1-Phosphatase loop with many substrates

The 1K1P loop can be expanded to include many substrates of the kinase and phosphatase. In this case we would have a system of enzymes such that:

$$
\begin{aligned}
E + S_1 &\underset{k_{-,E,1}}{\overset{k_{+,E,1}}{\rightleftharpoons}} ES_1 \xrightarrow{k_{cat,E,1}} E + S_1^* \\
E + S_2 &\underset{k_{-,E,2}}{\overset{k_{+,E,2}}{\rightleftharpoons}} ES_2 \xrightarrow{k_{cat,E,2}} E + S_2^* \\
&\vdots \\
E + S_N &\underset{k_{-,E,N}}{\overset{k_{+,E,N}}{\rightleftharpoons}} ES_N \xrightarrow{k_{cat,E,N}} E + S_N^*
\end{aligned}
$$

where $E = K$ or P. From these equations we have:

$$
[ES_1] = \frac{[E][S_1]}{K_{m,E,1}},\ [ES_2] = \frac{[E][S_2]}{K_{m,E,2}},\ \ldots,\ [ES_N] = \frac{[E][S_N]}{K_{m,E,N}} \tag{A.3.1}
$$

We also know from the conservation of mass of the enzyme:

$$
[E]_0 = [E] + [ES_1] + [ES_2] + \ldots + [ES_N] \tag{A.3.2}
$$

Substituting the system of equations from A.3.1 into A.3.2, we get:

$$
\begin{aligned}
[E]_0 &= [E]\left(1 + \frac{[S_1]}{K_{m,E,1}} + \frac{[S_2]}{K_{m,E,2}} + \ldots + \frac{[S_N]}{K_{m,E,N}}\right) \\
[E] &= \frac{[E]_0}{1 + \frac{[S_1]}{K_{m,E,1}} + \frac{[S_2]}{K_{m,E,2}} + \ldots + \frac{[S_N]}{K_{m,E,N}}} \\
[ES_1] &= \frac{[E]_0[S_1]}{[S_1] + K_{m,E,1}\left(1 + \frac{[S_2]}{K_{m,E,2}} + \ldots + \frac{[S_N]}{K_{m,E,N}}\right)}
\end{aligned} \tag{A.3.3}
$$

Substituting A.3.3 into the previously defined A.2.1.3, we arrive at:

$$\frac{d[S_1^*]}{dt} = \frac{V_{max,1}[S_1]}{\alpha K_{m,E,1} + [S_1]}$$

$$\alpha \equiv 1 + \sum_{i=2}^{N} \frac{[S_i]}{K_{m,E,i}}$$

From the above equation, we can proceed to solve for S_1^* as in section A.2.2; as expected, one obtains equation A.2.2.4, but with $\alpha_{K,1} \equiv 1 + \sum_{i=2}^{N}[S_i]/K_{m,K,i}$ and $\alpha_{P,1} \equiv 1 + \sum_{i=2}^{N}[S_i^*]/K_{m,P,i}$.

The increase in ultrasensitivity observed in Fig. 2.3*B* of the main text arises from the fact that, for the parameters we considered, at any $r_1 < 1$, the phosphatase has a higher maximum velocity than the kinase. As such, the majority of any substrates present will exist in the unphosphorylated form (i.e. $S_i^* < 0.5\ \forall i$). As more substrates are added, the accumulation of these unphosphorylated substrates begins to occupy the kinase, reducing free kinase concentration and thus reducing the "effective r" of the system. In the limit where N is large, the occupation increases until the kinase is completely saturated, ultimately leading to very low phosphorylation at $r_1 < 1$. For $r_1 > 1$, a similar situation holds, but with the phosphatase occupied by the S_i^*'s.

A.4 Analytical results for the 1-Kinase/2-Phosphatase loop

In this section we will show that S_1 phosphorylation always increases in $[S_2]_0$ in the limit in which $[S_1]_0 \ll K_m$. In this system S_1^* can be derived in a similar fashion to that for the 1K1P loop, resulting in:

$$S_1^* = \frac{(r_1-1) - (\alpha_{K,1}K_{K,1} + r_1 K_{P,1}) + \sqrt{((r_1-1) - (\alpha_{K,1}K_{K,1} + r_1 K_{P,1}))^2 + 4(r_1-1)r_1 K_{P,1}}}{2(r_1-1)} \tag{A.4.0.1}$$

Note this is similar to A.2.2.4, the difference being the absence of $\alpha_{P,1}$. This is because in this loop the substrates only share a kinase, making $\alpha_{P,1} = 1$. As such, $\frac{\partial S_1^*}{\partial \alpha_{P,1}} = 0$, by the chain rule we see:

$$\frac{dS_1^*}{d[S_2]_0} = \frac{dS_1^*}{d\alpha_{K,1}} \cdot \frac{d\alpha_{K,1}}{d[S_2]} \cdot \frac{d[S_2]}{d[S_2]_0} \tag{A.4.0.2}$$

Note that $\frac{dS_1^*}{d\alpha_{K,1}}$ is similar to $\frac{\partial S_1^*}{\partial \alpha_{K,1}}$ (A.2.3.4), the only difference being $\alpha_{P,1} = 1$ in this case. Since the value of $\alpha_{P,1}$ does not have an affect on the sign of $\frac{\partial S_1^*}{\partial \alpha_{K,1}}$, we can conclude that $\frac{dS_1^*}{d\alpha_{K,1}} < 0$ for any value of r_1 (see subsection A.2.3). Additionally, we can easily show $\frac{d\alpha_{K,1}}{d[S_2]} > 0$:

$$\begin{aligned} \alpha_{K,1} &= 1 + \frac{[S_2]}{K_{m,K,2}} \\ \frac{d\alpha_{K,1}}{d[S_2]} &= \frac{1}{K_{m,K,2}} > 0 \end{aligned} \tag{A.4.0.3}$$

A.4.1 $d[S_2]/d[S_2]_0$ is always positive

Using Mathematica [159], we can obtain an expression for $\frac{d[S_2]}{d[S_2]_0}$ at $r_2 \neq 1$. To simplify the derivation, we assume $[S_1]_0 \ll K_m$ so that $\alpha_{K,2} = 1$.

$$\begin{aligned} [S_2] &= (1 - S_2^*)[S_2]_0 \\ \frac{d[S_2]}{d[S_2]_0} &= 1 - S_2^* - \frac{dS_2^*}{d[S_2]_0}[S_2]_0 \\ &= 1 - \frac{-x' + \sqrt{(x')^2 + y'}}{2(r_2 - 1)} - \frac{z' + \frac{z'(-z') - \frac{y'}{2}}{\sqrt{(x')^2 + y'}}}{2(r_1 - 1)} \\ &= \frac{2(r_2 - 1) + x' - \sqrt{(x')^2 + y'} - z' + \frac{x'z' + \frac{y'}{2}}{\sqrt{(x')^2 + y'}}}{2(r_2 - 1)} \end{aligned} \tag{A.4.1.1}$$

In which:

$$x' \equiv -((r_2 - 1) - (K_{K,2} + r_2 K_{P,2})), \qquad y' \equiv 4(r_2 - 1) r_2 K_{P,2}, \qquad z' \equiv K_{K,2} + r_2 K_{P,2} \tag{A.4.1.2}$$

By the definitions of x' and z' we notice that $x' = -(r_2 - 1) + z'$, which can be substituted into A.4.1.1:

$$\begin{aligned}
\frac{d[S_2]}{d[S_2]_0} &= \frac{2(r_2-1) - (r_2-1) + z' - \sqrt{(x')^2 + y'} - z' + \frac{x'z' + \frac{y'}{2}}{\sqrt{(x')^2+y'}}}{2(r_2-1)} \\
&= \frac{(r_2-1) - \sqrt{(x')^2 + y'} + \frac{x'z' + \frac{y'}{2}}{\sqrt{(x')^2+y'}}}{2(r_2-1)} \\
&= \frac{(r_2-1)\sqrt{(x')^2+y'} - (x')^2 - y' + x'z' + \frac{y'}{2}}{2(r_2-1)\sqrt{(x')^2+y'}}
\end{aligned} \tag{A.4.1.3}$$

Additionally, by the definitions of x' and z', we see $(x')^2 = (r_2-1)^2 - 2(r_2-1)z' + (z')^2$ and $x'z' = -(r_2-1)z' + (z')^2$, which can be substituted into A.4.1.3:

$$\begin{aligned}
\frac{d[S_2]}{d[S_2]_0} &= \frac{(r_2-1)\sqrt{(x')^2+y'} - (r_2-1)^2 + 2(r_2-1)z' - (z')^2 - (r_2-1)z' + (z')^2 - \frac{y'}{2}}{2(r_2-1)\sqrt{(x')^2+y'}} \\
&= \frac{(r_2-1)\sqrt{(x')^2+y'} - (r_2-1)^2 + (r_2-1)z' - \frac{y'}{2}}{2(r_2-1)\sqrt{(x')^2+y'}} \\
&= \frac{\sqrt{(x')^2+y'} - (r_2-1) + z' - 2r_2K_{P,2}}{2\sqrt{(x')^2+y'}} \\
&= \frac{\sqrt{(x')^2+y'} + x' - 2r_2K_{P,2}}{2\sqrt{(x')^2+y'}}
\end{aligned} \tag{A.4.1.4}$$

We can show that $\frac{d[S_2]}{d[S_2]_0} > 0$ for all values of r_2 by assuming the opposite:

$$\begin{aligned}
\frac{d[S_2]}{d[S_2]_0} = \frac{\sqrt{(x')^2+y'} + x' - 2r_2K_{P,2}}{2\sqrt{(x')^2+y'}} &< 0 \\
\sqrt{(x')^2+y'} + x' - 2r_2K_{P,2} &< 0 \\
\sqrt{(x')^2+y'} &< -x' + 2r_2K_{P,2}
\end{aligned} \tag{A.4.1.5}$$

If the right hand side of A.4.1.5 is negative then we have already arrived at a contradiction. Otherwise we can square both sides without loss of information:

$$(x')^2 + y' < (x')^2 - 4r_2K_{P,2}x' + 4(r_2K_{P,2})^2 \tag{A.4.1.6}$$

$$y' < -4r_2K_{P,2}x' + 4(r_2K_{P,2})^2$$

$$4(r_2-1)r_2K_{P,2} < 4(r_2-1)r_2K_{P,2} - 4r_2K_{K,2}K_{P,2} - 4(r_2K_{P,2})^2 + 4(r_2K_{P,2})^2$$

$$0 < -4r_2K_{K,2}K_{P,2} \tag{A.4.1.7}$$

Which is clearly impossible, indicating $\frac{d[S_2]}{d[S_2]_0} > 0$ for $r_2 \neq 1$. Next we can obtain an expression for $\frac{d[S_2]}{d[S_2]_0}$ at $r_2 = 1$. At this point, S_2^* becomes:

$$S_2^* = \frac{K_{m,P,2}}{K_{m,K,2} + K_{m,P,2}} \tag{A.4.1.8}$$

As such, we can easily see that the derivative of A.4.1.8 with respect to $[S_2]_0$ is equal to zero. Applying this to the previous expression for $\frac{d[S_2]}{d[S_2]_0}$ (A.4.1.1) we notice that at $r_2 = 1$:

$$\frac{d[S_2]}{d[S_2]_0} = 1 - S_2^* \tag{A.4.1.9}$$

Since S_2^* must be a value between 0 and 1, it is easy to see that $\frac{d[S_2]}{d[S_2]_0} > 0$ at $r_2 = 1$, thus showing that $\frac{d[S_2]}{d[S_2]_0} > 0$ for all values of r_2.

A.4.2 $dS_1^*/d[S_2]_0$ is always negative

As previously shown, we can use the chain rule to define $\frac{dS_1^*}{d[S_2]_0}$ within this motif as:

$$\frac{dS_1^*}{d[S_2]_0} = \frac{dS_1^*}{d\alpha_{K,1}} \cdot \frac{d\alpha_{K,1}}{d[S_2]} \cdot \frac{d[S_2]}{d[S_2]_0} \tag{A.4.2.1}$$

In which $\frac{dS_1^*}{d\alpha_{K,1}} < 0$, $\frac{d\alpha_{K,1}}{d[S_2]} > 0$ and $\frac{d[S_2]}{d[S_2]_0} > 0$. Now we can see that $\frac{dS_1^*}{d[S_2]_0} < 0$ for all values of r_1 and r_2. At $r_2 < 1$, $\alpha_{K,1} > 1$ as most S_2 will be in the unphosphorylated form. Once $r_2 > 1$, S_2 switches

to its phosphorylated form, relieving the pressure on S_1 through $\alpha_{K,1}$, establishing the "gatekeeper" effect. We can see $\alpha_{K,1}$ approaches 1 as $r_2 \to \infty$, allowing S_1^* to behave as an isolated futile cycle in this limit. Since S_1^* is increasing in r_2, we can conclude that S_2 decreases S_1^* for all values of r_2 except in the limit $r_2 \to \infty$.

A.5 Analytical results for the 2-Kinase/1-Phosphatase loop

In this section we will show that S_1 phosphorylation also always increases in $[S_2]_0$ regardless of any other parameters. In this system S_1^* can be derived in a similar fashion to that for the 1K1P loop, resulting in:

$$S_1^* = \frac{(r_1-1)-(K_{K,1}+r_1\alpha_{P,1}K_{P,1})+\sqrt{((r_1-1)-(K_{K,1}+r_1\alpha_{P,1}K_{P,1}))^2+4(r_1-1)r_1\alpha_{P,1}K_{P,1}}}{2(r_1-1)} \tag{A.5.0.1}$$

Which is equivalent to A.2.2.4, the only difference being the lack of $\alpha_{K,1}$. As such $\frac{\partial S_1^*}{\partial \alpha_{K,1}} = 0$, and we notice that by the chain rule:

$$\frac{dS_1^*}{d[S_2]_0} = \frac{dS_1^*}{d\alpha_{P,1}} \cdot \frac{d\alpha_{P,1}}{d[S_2^*]} \cdot \frac{d[S_2^*]}{d[S_2]_0} \tag{A.5.0.2}$$

Note that $\frac{dS_1^*}{d\alpha_{P,1}}$ is similar to $\frac{\partial S_1^*}{\partial \alpha_{P,1}}$ (A.2.3.7), the only difference being $\alpha_{K,1} = 1$ in this case. Since the value of $\alpha_{K,1}$ does not have an affect on the sign of $\frac{\partial S_1^*}{\partial \alpha_{P,1}}$, we can conclude that $\frac{dS_1^*}{d\alpha_{P,1}} > 0$ for any value of r_1 (see subsection A.2.3). Additionally, we can easily show $\frac{d\alpha_{P,1}}{d[S_2]} > 0$:

$$\begin{aligned} \alpha_{P,1} &= 1 + \frac{[S_2^*]}{K_{m,P,2}} \\ \frac{d\alpha_{P,1}}{d[S_2]_0} &= \frac{1}{K_{m,P,2}} \end{aligned} \tag{A.5.0.3}$$

A.5.1 $d[S_2^*]/d[S_2]_0$ is always positive

We can define $[S_2^*]$ as:

$$[S_2^*] = S_2^*[S_2]_0 \tag{A.5.1.1}$$

And as such $\frac{d[S_2^*]}{d[S_2]_0}$ is:

$$\frac{d[S_2^*]}{d[S_2]_0} = S_2^* + \frac{dS_2^*}{d[S_2]_0}[S_2]_0 \tag{A.5.1.2}$$

Notice that $\frac{d[S_2^*]}{d[S_2]_0} = 1 - \frac{d[S_2]}{d[S_2]_0}$ (see A.4.1.1). We can then substitute A.4.1.4 in for $\frac{d[S_2]}{d[S_2]_0}$:

$$\begin{aligned}\frac{d[S_2^*]}{d[S_2]_0} &= 1 - \frac{\sqrt{(x')^2+y'}+x'-2r_2K_{P,2}}{2\sqrt{(x')^2+y'}} \\ &= \frac{\sqrt{(x')^2+y'}-x'+2r_2K_{P,2}}{2\sqrt{(x')^2+y'}}\end{aligned} \tag{A.5.1.3}$$

We can show $\frac{d[S_2^*]}{d[S_2]_0} > 0$ for any value of r_2 by assuming the opposite:

$$\begin{aligned}\frac{d[S_2^*]}{d[S_2]_0} = \frac{\sqrt{(x')^2+y'}-x'+2r_2K_{P,2}}{2\sqrt{(x')^2+y'}} &< 0 \\ \sqrt{(x')^2+y'}-x'+2r_2K_{P,2} &< 0 \\ \sqrt{(x')^2+y'} &< x'-2r_2K_{P,2}\end{aligned} \tag{A.5.1.4}$$

If the right hand side of A.5.1.4 is negative then we have already arrived at a contradiction. Otherwise we can square both sides without loss of information:

$$(x')^2+y' < (x')^2 - 4r_2K_{P,2}x' + 4(r_2K_{P,2})^2 \tag{A.5.1.5}$$

Note that this expression is the same as A.4.1.6, which we have already shown to be impossible, supporting the conclusion $\frac{d[S_2^*]}{d[S_2]_0} > 0$ for $r_2 \neq 0$. Next we can obtain an expression for $\frac{d[S_2]}{d[S_2]_0}$ at

$r_2 = 1$. At this point, S_2^* becomes:

$$S_2^* = \frac{K_{m,P,2}}{K_{m,K,2} + K_{m,P,2}} \tag{A.5.1.6}$$

As such, we can easily see that the derivative of A.5.1.6 with respect to $[S_2]_0$ is equal to zero. Applying this to the previous expression for $\frac{d[S_2^*]}{d[S_2]_0}$ (A.5.1.2) we notice that at $r_2 = 1$:

$$\frac{d[S_2^*]}{d[S_2]_0} = S_2^* \tag{A.5.1.7}$$

Since S_2^* must be a value between 0 and 1, it is easy to see that $\frac{d[S_2^*]}{d[S_2]_0} > 0$ at $r_2 = 1$, thus showing that $\frac{d[S_2^*]}{d[S_2]_0} > 0$ for all values of r_2.

A.5.2 $dS_1^*/d[S_2]_0$ is always positive

As previously shown, we can use the chain rule to define $\frac{dS_1^*}{d[S_2]_0}$ within this motif as:

$$\frac{dS_1^*}{d[S_2]_0} = \frac{dS_1^*}{d\alpha_{P,1}} \cdot \frac{d\alpha_{P,1}}{d[S_2^*]} \cdot \frac{d[S_2^*]}{d[S_2]_0} \tag{A.5.2.1}$$

In which $\frac{dS_1^*}{d\alpha_{P,1}} > 0$, $\frac{d\alpha_{P,1}}{d[S_2^*]} > 0$ and $\frac{d[S_2^*]}{d[S_2]_0} > 0$. Now we can see that $\frac{dS_1^*}{d[S_2]_0} > 0$ for all values of r_1 and r_2.

Appendix B

Appendix for Phosphatase Specificity and Pathway Insulation in Signaling Networks

B.1 Systems of Ordinary Differential Equations

B.1.1 2-Kinase/1-Phosphatase Loop with 2 Substrates

In order to characterize the effects of phosphatase saturation and competition on phosphatase-mediated crosstalk we used the 2-Kinase/1-Phosphatase Loop with 2 Substrates model that we have previously characterized [72]. The equations described below were derived and analyzed in our previous work; we include them here for completeness. The set of enzymatic reactions for the 2-Kinase/1-Phosphatase Loop with 2 Substrates model are:

$$K_1 + S_1 \underset{k_{-,K,1}}{\overset{k_{+,K,1}}{\rightleftharpoons}} K_1S_1 \xrightarrow{k_{cat,K,1}} K_1 + S_1^*$$

$$K_2 + S_2 \underset{k_{-,K,2}}{\overset{k_{+,K,2}}{\rightleftharpoons}} K_2S_2 \xrightarrow{k_{cat,K,2}} K_2 + S_2^*$$

$$P + S_1^* \underset{k_{-,P,1}}{\overset{k_{+,P,1}}{\rightleftharpoons}} PS_1^* \xrightarrow{k_{cat,P,1}} P + S_1$$

$$P + S_2^* \underset{k_{-,P,2}}{\overset{k_{+,P,2}}{\rightleftharpoons}} PS_2^* \xrightarrow{k_{cat,P,2}} P + S_2$$

Each contain three rates: the complex formation (k_+), the rate of complex dissociation (k_-), and catalytic rate (k_{cat}). These reactions are diagrammed in Fig. 3.2*A* of Chapter 3. The set of ODEs describing the free enzymes are:

$$\frac{d[K_1]}{dt} = [K_1S_1]k_{-,K,1} + [K_1S_1]k_{cat,K,1} - [K_1][S_1]k_{+,K,1}$$
$$\frac{d[K_2]}{dt} = [K_2S_2]k_{-,K,2} + [K_2S_2]k_{cat,K,2} - [K_2][S_2]k_{+,K,2}$$
$$\frac{d[P]}{dt} = [PS_1^*]k_{-,P,1} + [PS_2^*]k_{-,P,2} + [PS_1^*]k_{cat,P,1} + [PS_2^*]k_{cat,P,2} - [P][S_1^*]k_{+,P,1} - [P][S_2^*]k_{+,P,2}$$

The set of ODEs describing the free unphosphorylated substrates are:

$$\frac{d[S_1]}{dt} = [K_1S_1]k_{-,K,1} + [PS_1^*]k_{cat,P,1} - [K_1][S_1]k_{+,K,1}$$
$$\frac{d[S_2]}{dt} = [K_2S_2]k_{-,K,2} + [PS_2^*]k_{cat,P,2} - [K_2][S_2]k_{+,K,2}$$

The set of ODEs describing the free phosphorylated substrates are:

$$\frac{d[S_1^*]}{dt} = [PS_1^*]k_{-,P,1} + [K_1S_1]k_{cat,K,1} - [P][S_1^*]k_{+,P,1}$$
$$\frac{d[S_2^*]}{dt} = [PS_2^*]k_{-,P,2} + [K_2S_2]k_{cat,K,2} - [P][S_2^*]k_{+,P,2}$$

The set of ODEs describing the enzyme-substrate complexes are:

$$\frac{d[K_1S_1]}{dt} = [K_1][S_1]k_{+,K,1} - [K_1S_1]k_{-,K,1} - [K_1S_1]k_{cat,K,1}$$
$$\frac{d[K_2S_2]}{dt} = [K_2][S_2]k_{+,K,2} - [K_2S_2]k_{-,K,2} - [K_2S_2]k_{cat,K,2}$$
$$\frac{d[PS_1^*]}{dt} = [P][S_1^*]k_{+,P,1} - [PS_1^*]k_{-,P,1} - [PS_1^*]k_{cat,P,1}$$
$$\frac{d[PS_2^*]}{dt} = [P][S_2^*]k_{+,P,2} - [PS_2^*]k_{-,P,2} - [PS_2^*]k_{cat,P,2}$$

For purposes of display we used the following rate parameters:

Parameter	Value
$k_{+,K,i}$	0.001 nM^{-1}s^{-1}
$k_{-,K,i}$	0.1-999.1 s^{-1}
$k_{cat,K,i}$	0.9 s^{-1}
$k_{+,P,i}$	0.001 nM^{-1}s^{-1}
$k_{-,P,i}$	0.1-999.1 s^{-1}
$k_{cat,P,i}$	0.9 s^{-1}

where $i = 1$ or 2. The ranges listed for the dissociation rates (i.e. $k_{-,K,i}$) are used to set the K_M's of the enzymes in different simulations. Note that, while the values of these parameters are not meant to describe any specific enzyme, they are within the range of values obtained for kinases and phosphatases experimentally [96, 97, 160].

We used the following initial conditions for all of our simulations:

Molecular Species	Initial Concentration
K_1	0-20 nM
K_2	0-20 nM
P	10 nM, 1μM
S_1	10 μM
S_2	0,10 μM

with the remaining molecular species having initial concentrations of 0. The ranges of concentrations of K_1 and K_2 are used to vary the values of r_1 and r_2.

This model was used to generate Fig. 3.1*B* and *C* of Chapter 3. The concentration of K_2 was set to 0 for $r_2 = 0$ and to 20 nM for $r_2 = 2$. The concentration of K_1 was set between 0-20 nM to vary r_1 between 0 and 2. Both substrates are present at a concentration of 10μM. The values of $k_{-,K,1}$ and $k_{-,P,1}$ were both set to 999.1 s^{-1} so that $K_{M,K,1} = K_{M,P,1} = 100 \times [S_1]_0$, while $k_{-,K,2}$ and $k_{-,P,2}$ were set to 0.1 s^{-1} so that $K_{M,P,2} = K_{M,K,2} = 100 \times [S_2]_0$.

We also used this model to generate Fig. 3.2 of Chapter 3. In Fig. 3.3*A*, $K_{M,P,i}$ was set by changing the values of $k_{-,P,1}$ and $k_{-,P,2}$ between 0.1-999.1 s^{-1}. The value of r_1 was set in Fig.

3.2*B* by setting the concentration of K_1 between 0-20 nM, and the concentration of K_2 was set to 0 for $r_2 = 0$ and to 20 nM for $r_2 = 2$. For Fig. 3.2*C*, we first ran the model with $K_2 = S_2 = 0$ to steady state with the initial concentration of K_1 at 20 nM ($r_1 = 2$). We then removed all K_1 molecules from the system. The bound S_1 was added back to the concentration of unphosphorylated S_1. The simulations were then resumed to obtain the time courses visualized in Fig. 3.2*C*. $K_{M,P}$ was set by using values of $k_{-,P,2,1}$ = 0.1 s^{-1}, 9.1 s^{-1}, and 99.1 s^{-1}. The fraction of phosphorylated S_1^* was normalized so that $\hat{S}_1^*(t) = (max(S_1^*) - S_1^*(t))/(max(S_1^*) - min(S_1^*))$. Fig. 3.2*D* was obtained using the same procedures as in Fig. 3.2*C*, setting $K_{M,P}$ by using values of $k_{-,P,2,1}$ = 0.1-999.1 s^{-1}. The half-time of S_1^* phosphorylation was obtained by finding the time $t_{1/2} = t$ where $\hat{S}_1^*(t)$ = 0.5. The total concentration of the phosphatase was set to either 10 nM (magenta curve), or 1 μM (purple curve).

To test the effectiveness of an increased phosphatase concentration in insulating substrates against phosphatase crosstalk while maintaining strong $K_{M,P,i}$ values, we set $K_{M,P,i} = 1\mu$M, $r_1 = 0.05$, $r_2 = 2$ and varied the concentration of the phosphatase from 10 nM to 1 mM (Fig. B.1).

B.1.2 2-Kinase/1-Phosphatase Loop with 2 Substrates and 2 Reservoir Proteins

In order to characterize the effects of reservoir proteins that bind to and shield phosphorylated substrates from dephosphorylated on phosphatase-mediated crosstalk we expanded the 2-Kinase/1-Phosphatase Loop with 2 Substrates model to include two substrate-specific reservoir proteins, R_1

and R_2. The set of enzymatic reactions for the model are:

$$K_1 + S_1 \underset{k_{-,K,1}}{\overset{k_{+,K,1}}{\rightleftharpoons}} K_1S_1 \xrightarrow{k_{cat,K,1}} K_1 + S_1^*$$

$$K_2 + S_2 \underset{k_{-,K,2}}{\overset{k_{+,K,2}}{\rightleftharpoons}} K_2S_2 \xrightarrow{k_{cat,K,2}} K_2 + S_2^*$$

$$P + S_1^* \underset{k_{-,P,1}}{\overset{k_{+,P,1}}{\rightleftharpoons}} PS_1^* \xrightarrow{k_{cat,P,1}} P + S_1$$

$$P + S_2^* \underset{k_{-,P,2}}{\overset{k_{+,P,2}}{\rightleftharpoons}} PS_2^* \xrightarrow{k_{cat,P,2}} P + S_2$$

$$R_1 + S_1^* \underset{k_{-,R,1}}{\overset{k_{+,R,1}}{\rightleftharpoons}} R_1S_1^*$$

$$R_2 + S_2^* \underset{k_{-,R,2}}{\overset{k_{+,R,2}}{\rightleftharpoons}} R_2S_2^*$$

The set of ODEs describing the free enzymes are:

$$\frac{d[K_1]}{dt} = [K_1S_1]k_{-,K,1} + [K_1S_1]k_{cat,K,1} - [K_1][S_1]k_{+,K,1}$$
$$\frac{d[K_2]}{dt} = [K_2S_2]k_{-,K,2} + [K_2S_2]k_{cat,K,2} - [K_2][S_2]k_{+,K,2}$$
$$\frac{d[P]}{dt} = [PS_1^*]k_{-,P,1} + [PS_2^*]k_{-,P,2} + [PS_1^*]k_{cat,P,1} + [PS_2^*]k_{cat,P,2} - [P][S_1^*]k_{+,P,1} - [P][S_2^*]k_{+,P,2}$$

The set of ODEs describing the free unphosphorylated substrates are:

$$\frac{d[S_1]}{dt} = [K_1S_1]k_{-,K,1} + [PS_1^*]k_{cat,P,1} - [K_1][S_1]k_{+,K,1}$$
$$\frac{d[S_2]}{dt} = [K_2S_2]k_{-,K,2} + [PS_2^*]k_{cat,P,2} - [K_2][S_2]k_{+,K,2}$$

The set of ODEs describing the free phosphorylated substrates are:

$$\frac{d[S_1^*]}{dt} = [PS_1^*]k_{-,P,1} + [K_1S_1]k_{cat,K,1} + [R_1S_1^*]k_{-,R,1} - [P][S_1^*]k_{+,P,1} - [R_1][S_1^*]k_{+,R,1}$$
$$\frac{d[S_2^*]}{dt} = [PS_2^*]k_{-,P,2} + [K_2S_2]k_{cat,K,2} + [R_2S_2^*]k_{-,R,2} - [P][S_2^*]k_{+,P,2} - [R_2][S_2^*]k_{+,R,2}$$

The set of ODEs describing the enzyme-substrate complexes are:

$$\frac{d[K_1S_1]}{dt} = [K_1][S_1]k_{+,K,1} - [K_1S_1]k_{-,K,1} - [K_1S_1]k_{cat,K,1}$$
$$\frac{d[K_2S_2]}{dt} = [K_2][S_2]k_{+,K,2} - [K_2S_2]k_{-,K,2} - [K_2S_2]k_{cat,K,2}$$
$$\frac{d[PS_1^*]}{dt} = [P][S_1^*]k_{+,P,1} - [PS_1^*]k_{-,P,1} - [PS_1^*]k_{cat,P,1}$$
$$\frac{d[PS_2^*]}{dt} = [P][S_2^*]k_{+,P,2} - [PS_2^*]k_{-,P,2} - [PS_2^*]k_{cat,P,2}$$

The set of ODEs describing the free reservoir proteins are:

$$\frac{d[R_1]}{dt} = [R_1S_1^*]k_{-,R,1} - [R_1][S_1^*]k_{+,R,1}$$
$$\frac{d[R_2]}{dt} = [R_2S_2^*]k_{-,R,2} - [R_2][S_2^*]k_{+,R,2}$$

The set of ODEs describing the reservoir-substrate complexes are:

$$\frac{d[R_1S_1^*]}{dt} = [R_1][S_1^*]k_{+,R,1} - [R_1S_1^*]k_{-,R,1}$$
$$\frac{d[R_2S_2^*]}{dt} = [R_2][S_2^*]k_{+,R,2} - [R_2S_2^*]k_{-,R,2}$$

For purposes of display we used the following rate parameters:

Parameter	Value
$k_{+,K,i}$	0.001 $\text{nM}^{-1}\text{s}^{-1}$
$k_{-,K,i}$	0.1-999.1 s^{-1}
$k_{cat,K,i}$	0.9 s^{-1}
$k_{+,P,i}$	0.001 $\text{nM}^{-1}\text{s}^{-1}$
$k_{-,P,i}$	0.1-999.1 s^{-1}
$k_{cat,P,i}$	0.9 s^{-1}
$k_{+,R,i}$	0.001 $\text{nM}^{-1}\text{s}^{-1}$
$k_{-,R,i}$	0.01 s^{-1}

where $i = 1$ or 2. The ranges listed for the dissociation rates (i.e. $k_{-,K,i}$) are used to set the K_M's of the enzymes in different simulations. Note that, while the values of these parameters are not meant to describe any specific enzyme, they are within the range of values obtained for kinases and phosphatases experimentally [96, 97, 160].

We used the following initial conditions for all of our simulations:

Molecular Species	Initial Concentration
K_1	0-20 nM
K_2	0-20 nM
P	10 nM, 1μM
S_1	10 μM
S_2	0,10 μM
R_1	0-100 μM
R_2	0-100 μM

with the remaining molecular species having initial concentrations of 0. The ranges of concentrations of K_1 and K_2 are used to vary the values of r_1 and r_2.

B.1.3 2-Kinase/1-Phosphatase Loop with Many Substrates

The expression for the fraction of phosphorylation S_1^* in a 2K1P loop in which kinase K_2 and P act upon N substrates is similar to the expression for a 2K1P loop with 2 substrates we have previously derived (equation 1 in the main text) [72]. The only difference in this case is the value of $\alpha_{P,1}$, the inhibitory term that captures the effects of the competing substrates on the phosphatase. In a model with N substrates, this term becomes:

$$\alpha_{P,1} = 1 + \sum_{i \neq 1} \frac{[S_i^*]}{K_{M,P,i}} \tag{B.1.3.1}$$

Note that the value of $\alpha_{P,1}$, and thus the inhibition of the phosphatase, depends on the total saturation of the phosphatase across all its substrates. A set of substrates that all respond to the same signal can thus cause phosphatase crosstalk with other proteins in the network, even if none of those substrates is at high enough concentration to saturate the phosphatase individually.

B.1.4 1-Kinase/1-Substrate Model with Synthesis and Degradation

In order to characterize the effectiveness of synthesis and degradation of a substrate as a replacement for the phosphatase, we created the 1-Kinase/1-Substrate model with Synthesis and Degradation. The set of enzymatic reactions for this system are:

$$K + S \underset{k_-}{\overset{k_+}{\rightleftharpoons}} KS \xrightarrow{k_{cat}} K + S^*$$

$$S \xrightarrow{k_{deg,U}} \emptyset$$

$$KS \xrightarrow{k_{deg,U}} K$$

$$S^* \xrightarrow{k_{deg,P}} \emptyset$$

$$\emptyset \xrightarrow{k_{synth}} S$$

where $k_{deg,U}$ and $k_{deg,P}$ are the degradation rates of unphosphorylated and phosphorylated S. This model includes separate degradation rates for the unphosphorylated and phosphorylated substrate ($k_{deg,U}$ and $k_{deg,P}$), since phosphorylation of the substrate might either increase or decrease the stability of the protein. The ODE describing the free kinase is:

$$\frac{d[K]}{dt} = [KS]k_- + [KS]k_{cat} + [KS]k_{deg,U} - [K][S]k_+$$

The ODE describing the free unphosphorylated substrate is:

$$\frac{d[S]}{dt} = [KS]k_- + k_{synth} - [K][S]k_+ - [S]k_{deg,U} \tag{B.1.4.1}$$

The ODE describing the free phosphorylated substrate is:

$$\frac{d[S^*]}{dt} = [KS]k_{cat} - [S^*]k_{deg,P}$$

The ODE describing the concentration of kinase-substrate complex is:

$$\frac{d[KS]}{dt} = [K][S]k_+ - [KS]k_- - [KS]k_{cat} - [KS]k_{deg,U}$$

In order to understand the effects of degradation at steady state, note that the total substrate concentration is defined as:

$$[S]_T = [S] + [S^*] + [KS]$$

and the change in total substrate concentration can thus be written:

$$\frac{d[S]_T}{dt} = \frac{d[S]}{dt} + \frac{d[S^*]}{dt} + \frac{d[KS]}{dt}$$

At steady state, $d[S]_T/dt = 0$. By substituting the above ODEs and simplifying, we get:

$$\frac{d[S]_T}{dt} = k_{synth} - ([S] + [KS])k_{deg,U} - [S^*]k_{deg,P} = 0$$

We can then solve for k_{synth}:

$$k_{synth} = ([S] + [KS])k_{deg,U} + [S^*]k_{deg,P}$$

We can substitute this equation into the original differential equation for $[S]$ (equation B.1.4.1):

$$\frac{d[S]}{dt} = [KS]k_- - [K][S]k_+ + [KS]k_{deg,U} + [S^*]k_{deg,P} \tag{B.1.4.2}$$

Equation B.1.4.2 is useful for two reasons. For one, there is a positive term in the equation corresponding to the degradation of phosphorylated substrate ($[S^*]k_{deg,P}$). This term reflects the fact that, in order for $[S]_T$ to remain constant at steady state, new, unphosphorylated substrate molecules must be synthesized to replace S^* molecules that are degraded. There is thus an "effective" dephosphorylation rate in this system where S^* molecules are converted to S, which corresponds mathematically to an unsaturateable first-order phosphatase. Secondly, we used equation B.1.4.2 instead of B.1.4.1 in our numerical integration, so $[S]_T = [S]_0$ for all time; in other words, while the concentration of unphosphorylated and phosphorylated substrate might change in our simulations, the total concentration of substrate remains constant. This allows us to control total substrate levels by setting the initial substrate concentration, as we do in our other models. One could of course simulate equation B.1.4.1 with a constant k_{synth} that allows total substrate concentration to vary with time; while such transients might have interesting effects on the system, we leave consideration of those effects to future work.

For purposes of display we used the following rate parameters:

Parameter	Value
k_+	0.001 $nM^{-1}s^{-1}$
k_-	0.1 s^{-1}
k_{cat}	0.9 s^{-1}
$k_{deg,U}$	1x10^{-7} - 1x10^{-4} s^{-1}
$k_{deg,P}$	1x10^{-7} - 1x10^{-4} s^{-1}

The ranges of $k_{deg,U}$ and $k_{deg,P}$ were used to vary the degradation rate of the substrate across simulations. The simulations started with the following initial concentrations:

Molecular Species	Initial Concentration
K	0 - 1 nM
S	10 μM
KS	0

The range of K was used to set the value of r_{deg} across simulations.

This model was used to generate Fig. 3.3*B* and *C* of Chapter 3. In Fig. 3.3*B* we used $k_{deg,U} = k_{deg,P} = \log 2/t_{1/2}$ for $t_{1/2} = 10$, 31, and 187 hrs. The value of r_{deg} was set between 0 and 2 by changing the initial concentration of K so that $[K] = (r_{deg}[S]_0 k_{deg,P})/k_{cat}$ (where r_{deg} was set to the desired value, see Section B.3 for the derivation of r_{deg}), which ends up giving a value between 0-0.186 nM. In Fig. 3.3*C* we first ran the model to steady state with an initial concentration of K at 0.186 nM ($r_{deg} = 2$). All kinase molecules were then removed from the system, and the simulation continued in order to obtain the time courses shown in Fig. 3.3*C*.

B.1.5 2-Kinase/1-Phosphatase Loop with “Ordered" Phosphatase Adaptors

In order to characterize the effectiveness of phosphatase adaptors in insulating pathways from phosphatase crosstalk, we first developed the 2-Kinase/1-Phosphatase Loop with Ordered Phos-

phatase Adaptors. The set of enzymatic reactions for this model are:

$$K_1 + S_1 \underset{k_{-,K,1}}{\overset{k_{+,K,1}}{\rightleftharpoons}} K_1S_1 \xrightarrow{k_{cat,K,1}} K_1 + S_1^*$$
$$K_2 + S_2 \underset{k_{-,K,2}}{\overset{k_{+,K,2}}{\rightleftharpoons}} K_2S_2 \xrightarrow{k_{cat,K,2}} K_2 + S_2^*$$
$$PA_1 + S_1^* \underset{k_{-,P,1}}{\overset{k_{+,P,1}}{\rightleftharpoons}} PA_1S_1^* \xrightarrow{k_{cat,P,1}} PA_1 + S_1$$
$$PA_2 + S_2^* \underset{k_{-,P,2}}{\overset{k_{+,P,2}}{\rightleftharpoons}} PA_2S_2^* \xrightarrow{k_{cat,P,2}} PA_2 + S_2$$

$$P + A_1 \underset{k_{-,A,1}}{\overset{k_{+,A,1}}{\rightleftharpoons}} PA_1$$
$$P + A_2 \underset{k_{-,A,2}}{\overset{k_{+,A,2}}{\rightleftharpoons}} PA_2$$

where $k_{-,A,i}$ represents the dissociation rate of the phosphatase-adaptor complex and $k_{+,A,i}$ represents the association rate. The reactions that include a phosphatase molecule are diagrammed in the inset of Fig. 3.5*A* of Chapter 3. The set of ODEs describing the concentration of free enzymes are:

$$\frac{d[K_1]}{dt} = [K_1S_1]k_{-,K,1} + [K_1S_1]k_{cat,K,1} - [K_1][S_1]k_{+,K,1}$$
$$\frac{d[K_2]}{dt} = [K_2S_2]k_{-,K,2} + [K_2S_2]k_{cat,K,2} - [K_2][S_2]k_{+,K,2}$$
$$\frac{d[P]}{dt} = [PA_1]k_{-,A,1} + [PA_2]k_{-,A,2} - [P][A_1]k_{+,A,1} - [P][A_2]k_{+,A,2}$$

The set of ODEs describing the concentration of free unphosphorylated substrates are:

$$\frac{d[S_1]}{dt} = [K_1S_1]k_{-,K,1} + [PA_1S_1^*]k_{cat,P,1} - [K_1][S_1]k_{+,K,1}$$
$$\frac{d[S_2]}{dt} = [K_2S_2]k_{-,K,2} + [PA_2S_2^*]k_{cat,P,2} - [K_2][S_2]k_{+,K,2}$$

The set of ODEs describing the concentration of free phosphorylated substrates are:

$$\frac{d[S_1^*]}{dt} = [PA_1S_1^*]k_{-,P,1} + [K_1S_1]k_{cat,K,1} - [PA_1][S_1^*]k_{+,P,1}$$
$$\frac{d[S_2^*]}{dt} = [PA_2S_2^*]k_{-,P,2} + [K_2S_2]k_{cat,K,2} - [PA_2][S_2^*]k_{+,P,2}$$

The set of ODEs describing the concentration of adaptor-bound phosphatase are:

$$\frac{d[PA_1]}{dt} = [P][A_1]k_{+,A,1} + [PA_1S_1^*]k_{-,P,1} + [PA_1S_1^*]k_{cat,P,1} - [PA_1]k_{-,A,1} - [PA_1][S_1^*]k_{+,P,1}$$
$$\frac{d[PA_2]}{dt} = [P][A_2]k_{+,A,2} + [PA_2S_2^*]k_{-,P,2} + [PA_2S_2^*]k_{cat,P,2} - [PA_2]k_{-,A,2} - [PA_2][S_2^*]k_{+,P,2}$$

The set of ODEs describing the concentration of enzyme-substrate complexes are:

$$\frac{d[K_1S_1]}{dt} = [K_1][S_1]k_{+,K,1} - [K_1S_1]k_{-,K,1} - [K_1S_1]k_{cat,K,1}$$
$$\frac{d[K_2S_2]}{dt} = [K_2][S_2]k_{+,K,2} - [K_2S_2]k_{-,K,2} - [K_2S_2]k_{cat,K,2}$$
$$\frac{d[PA_1S_1^*]}{dt} = [PA_1][S_1^*]k_{+,P,1} - [PA_1S_1^*]k_{-,P,1} - [PA_1S_1^*]k_{cat,P,1}$$
$$\frac{d[PA_2S_2^*]}{dt} = [PA_2][S_2^*]k_{+,P,2} - [PA_2S_2^*]k_{-,P,2} - [PA_2S_2^*]k_{cat,P,2}$$

For purposes of display we used the following rate parameters:

Parameter	Value
$k_{+,K,i}$	0.001 $\text{nM}^{-1}\text{s}^{-1}$
$k_{-,K,i}$	0.1 s^{-1}
$k_{cat,K,i}$	0.9 s^{-1}
$k_{+,P,i}$	0.001 $\text{nM}^{-1}\text{s}^{-1}$
$k_{-,P,i}$	0.1 s^{-1}
$k_{cat,P,i}$	0.9 s^{-1}
$k_{+,A,i}$	0.001 $\text{nM}^{-1}\text{s}^{-1}$
$k_{-,A,i}$	0.1 s^{-1}

where $i = 1$ or 2. Our simulations started with the following initial concentrations:

Molecular Species	Initial Concentration
K_1	0-20 nM
K_2	0-20 nM
P	10 nM
S_1	10 μM
S_2	10 μM
A_1	10^{-1}-10^4 nM
A_2	10^{-1}-10^4 nM

This model is used to generate Fig. 3.4 of Chapter 3. In Fig. 3.4*A*, r_1 is set to 0.1 by having an initial concentration of K_1 of 1 nM and r_2 is set to 2 by having an initial concentration of K_2 of 20 nM. The initial concentration of the adaptors A_1 and A_2 were then concurrently varied between 10^{-1} to 10^4 nM. In Fig. 3.4*B*, A_1 and A_2 were initialized with total concentration of 10 nM each. The values of r_1 and r_2 were set between 0 and 2 by setting the initial concentrations of K_1 and K_2 between 0 and 20 nM.

B.1.6 2-Kinase/1-Phosphatase Loop with “Unordered" Phosphatase Adaptors

We also developed a model of a 2-Kinase/1-Phosphatase Loop with Unordered Phosphatase Adaptors. The set of enzymatic reactions for this model are:

$$K_1 + S_1 \underset{k_{-,K,1}}{\overset{k_{+,K,1}}{\rightleftharpoons}} K_1S_1 \xrightarrow{k_{cat,K,1}} K_1 + S_1^*$$
$$K_2 + S_2 \underset{k_{-,K,2}}{\overset{k_{+,K,2}}{\rightleftharpoons}} K_2S_2 \xrightarrow{k_{cat,K,2}} K_2 + S_2^*$$

$$PA_1 + S_1^* \underset{k_{-,P,1}}{\overset{k_{+,P,1}}{\rightleftharpoons}} PA_1S_1^*$$
$$PA_2 + S_2^* \underset{k_{-,P,2}}{\overset{k_{+,P,2}}{\rightleftharpoons}} PA_2S_2^*$$
$$P + A_1S_1^* \underset{k_{-,A,1}}{\overset{k_{+,A,1}}{\rightleftharpoons}} PA_1S_1^*$$
$$P + A_2S_2^* \underset{k_{-,A,2}}{\overset{k_{+,A,2}}{\rightleftharpoons}} PA_2S_2^*$$
$$PA_1S_1^* \underset{k_{+,P,1}}{\overset{k_{cat,P,1}}{\rightleftharpoons}} PA_1 + S_1^*$$
$$PA_2S_2^* \underset{k_{+,P,2}}{\overset{k_{cat,P,2}}{\rightleftharpoons}} PA_2 + S_2^*$$

$$P + A_1 \underset{k_{-,A,1}}{\overset{k_{+,A,1}}{\rightleftharpoons}} PA_1$$
$$P + A_2 \underset{k_{-,A,2}}{\overset{k_{+,A,2}}{\rightleftharpoons}} PA_2$$
$$A_1 + S_1^* \underset{k_{-,P,1}}{\overset{k_{+,P,1}}{\rightleftharpoons}} A_1S_1^*$$
$$A_2 + S_2^* \underset{k_{-,P,2}}{\overset{k_{+,P,2}}{\rightleftharpoons}} A_2S_2^*$$

(B.1.6.1)

This model differs from the “ordered" model in that the adaptors can first bind the phosphorylated substrate without needing to be bound to a phosphatase catalytic core. These reactions are diagrammed in the inset of Fig. 3.5*A* of Chapter 3. The set of ODEs describing the concentration of free enzymes are:

$$\frac{d[K_1]}{dt} = [K_1S_1]k_{-,K,1} + [K_1S_1]k_{cat,K,1} - [K_1][S_1]k_{+,K,1}$$
$$\frac{d[K_2]}{dt} = [K_2S_2]k_{-,K,2} + [K_2S_2]k_{cat,K,2} - [K_2][S_2]k_{+,K,2}$$

$$\frac{d[P]}{dt} = [PA_1]k_{-,A,1} + [PA_2]k_{-,A,2} + [PA_1S_1^*]k_{-,A,1} + [PA_2S_2^*]k_{-,A,2}$$
$$- [P][A_1]k_{+,A,1} - [P][A_2]k_{+,A,2} - [P][A_1S_1^*]k_{+,A,1} - [P][A_2S_2^*]k_{+,A,2}$$

The set of ODEs describing the concentration of free unphosphorylated substrates are:

$$\frac{d[S_1]}{dt} = [K_1S_1]k_{-,K,1} + [PA_1S_1^*]k_{cat,P,1} - [K_1][S_1]k_{+,K,1}$$
$$\frac{d[S_2]}{dt} = [K_2S_2]k_{-,K,2} + [PA_2S_2^*]k_{cat,P,2} - [K_2][S_2]k_{+,K,2}$$

The set of ODEs describing the concentration of free phosphorylated substrates are:

$$\frac{d[S_1^*]}{dt} = [PA_1S_1^*]k_{-,P,1} + [A_1S_1^*]k_{-,P,1} + [K_1S_1]k_{cat,K,1} - [PA_1][S_1^*]k_{+,P,1} - [A_1][S_1^*]k_{+,P,1}$$
$$\frac{d[S_2^*]}{dt} = [PA_2S_2^*]k_{-,P,2} + [A_2S_2^*]k_{-,P,2} + [K_2S_2]k_{cat,K,2} - [PA_2][S_2^*]k_{+,P,2} - [A_2][S_2^*]k_{+,P,2}$$

The set of ODEs describing the concentration of adaptor-bound phosphatase are:

$$\frac{d[PA_1]}{dt} = [P][A_1]k_{+,A,1} + [PA_1S_1^*]k_{-,P,1} + [PA_1S_1^*]k_{cat,P,1} - [PA_1]k_{-,A,1} - [PA_1][S_1^*]k_{+,P,1}$$
$$\frac{d[PA_2]}{dt} = [P][A_2]k_{+,A,2} + [PA_2S_2^*]k_{-,P,2} + [PA_2S_2^*]k_{cat,P,2} - [PA_2]k_{-,A,2} - [PA_2][S_2^*]k_{+,P,2}$$

The set of ODEs describing the concentration of adaptor-bound phosphorylated substrate are:

$$\frac{d[A_1S_1^*]}{dt} = [A_1][S_1^*]k_{+,P,1} + [PA_1S_1^*]k_{-,A,1} - [A_1S_1^*]k_{-,P,1} - [P][A_1S_1^*]k_{+,A,1}$$
$$\frac{d[A_2S_2^*]}{dt} = [A_2][S_2^*]k_{+,P,2} + [PA_2S_2^*]k_{-,A,2} - [A_2S_2^*]k_{-,P,2} - [P][A_2S_2^*]k_{+,A,2}$$

The set of ODEs describing the concentration of enzyme-substrate complexes are:

$$\frac{d[K_1S_1]}{dt} = [K_1][S_1]k_{+,K,1} - [K_1S_1]k_{-,K,1} - [K_1S_1]k_{cat,K,1}$$
$$\frac{d[K_2S_2]}{dt} = [K_2][S_2]k_{+,K,2} - [K_2S_2]k_{-,K,2} - [K_2S_2]k_{cat,K,2}$$
$$\frac{d[PA_1S_1^*]}{dt} = [PA_1][S_1^*]k_{+,P,1} + [P][A_1S_1^*]k_{+,P,1} - [PA_1S_1^*]k_{-,P,1} - [PA_1S_1^*]k_{-,A,1} - [PA_1S_1^*]k_{cat,P,1}$$
$$\frac{d[PA_2S_2^*]}{dt} = [PA_2][S_2^*]k_{+,P,2} + [P][A_2S_1^*]k_{+,P,2} - [PA_2S_2^*]k_{-,P,2} - [PA_2S_2^*]k_{-,A,2} - [PA_2S_2^*]k_{cat,P,2}$$

For purposes of display we used the following rate parameters:

Parameter	Value
$k_{+,K,i}$	0.001 nM^{-1}s^{-1}
$k_{-,K,i}$	0.1 s^{-1}
$k_{cat,K,i}$	0.9 s^{-1}
$k_{+,P,i}$	0.001 nM^{-1}s^{-1}
$k_{-,P,i}$	0.1 s^{-1}
$k_{cat,P,i}$	0.9 s^{-1}
$k_{+,A,i}$	0.001 nM^{-1}s^{-1}
$k_{-,A,i}$	0.1 s^{-1}

where $i = 1$ or 2. We set the $K_{D,A,i} = k_{-,A,i}/k_{+,A,i} = 100$ nM to represent a reasonably high affinity of the phosphatase catalytic core P with the adaptor domains A_i. We used the same affinity for the binding of the adaptor domain to the substrate. Note that, in this model, the affinity of the phosphatase for the adaptor domain, and the affinity of the adaptor domain for the substrate, does not depend on wether the adaptor is bound to its other partner. Our simulations started with the following initial concentrations:

Molecular Species	Initial Concentration
K_1	0-20 nM
K_2	0-20 nM
P	10 nM
S_1	10 μM
S_2	0-10 μM
A_1	10^{-1}-10^4 nM
A_2	0-10^4 nM

This model is used to generate Fig. 3.5 of Chapter 3. In Fig. 3.5*A*, r_1 is set to 0.1 by having an initial concentration of K_1 of 1 nM and r_2 is set to 2 by having an initial concentration of K_2 of 20 nM. The initial concentration of the adaptors A_1 and A_2 were then concurrently varied between

10^{-1} to 10^4 nM. In Fig. 3.5*B*, A_1 and A_2 were initialized with total concentration of 10 nM each. The values of r_1 and r_2 were set between 0 and 2 by setting the initial concentration of K_1 and K_2 to between 0 and 20 nM.

B.2 The responses of the substrates of an unsaturated phosphatase are strictly hyperbolic in r

We have previously shown [72], following the derivation of Goldbeter and Koshland [26], that the change in product concentration S_1^* for an enzyme E with multiple substrates can be defined as:

$$\frac{d[S_1^*]}{dt} = \frac{V_{max,E,1}[S_1]}{\alpha_{E,1}K_{M,E,1} + [S_1]} \tag{B.2.0.1}$$

where $\alpha_{E,1} \equiv 1 + \sum_{i \neq 1} \frac{[S_i]}{K_{M,E,i}}$ is an inhibitory constant for substrate competition with S_1 for E. For a futile cycle at steady state we will have $d[S_1^*]/dt = d[S_1]/dt$. For an unsaturated 2K1P loop, $\alpha_{P,i} = 1$ since $K_{M,P,i} \gg [S_i]_0$. Given B.2.0.1, at steady state we have:

$$\frac{V_{max,K,1}[S_1]}{K_{m,K,1} + [S_1]} = \frac{V_{max,P,1}[S_1^*]}{K_{M,P,1} + [S_1^*]} \tag{B.2.0.2}$$

Following the standard Michaelis-Menten approach [26, 72], we assume that $[S_i]_0 \gg [K]_0, [P]_0$, giving us $[S_i]_0 = [S_i] + [S_i^*]$, which can be substituted into B.2.0.2:

$$\frac{V_{max,K,1}(1 - S_1^*)}{K_{K,1} + 1 - S_1^*} = \frac{V_{max,P,1}S_1^*}{K_{P,1} + S_1^*} \tag{B.2.0.3}$$

where $S_1 \equiv [S_1]/[S_1]_0$, $S_1^* \equiv [S_1^*]/[S_1]_0$, $K_{K,1} \equiv K_{M,K,1}/[S_1]_0$, and $K_{P,1} \equiv K_{M,P,1}/[S_1]_0$. Dividing both sides by $V_{max,P,1}$ we obtain:

$$\frac{r_1(1 - S_1^*)}{K_{K,1} + 1 - S_1^*} = \frac{S_1^*}{K_{P,1} + S_1^*} \tag{B.2.0.4}$$

where $r_1 \equiv V_{max,K,1}/V_{max,P,1}$. Since we are assuming the phosphatase to be unsaturated, $K_{P,1} \gg S_1^*$, and as such B.2.0.4 can be simplified to:

$$\frac{r_1(1-S_1^*)}{K_{K,1}+1-S_1^*} = \frac{S_1^*}{K_{P,1}} \tag{B.2.0.5}$$

Expanding B.2.0.5, we get:

$$r_1 K_{P,1} - r_1 K_{P,1} S_1^* = K_{K,1} S_1^* + S_1^* - (S_1^*)^2$$

$$(S_1^*)^2 - (1 + K_{K,1} + r_1 K_{P,1}) S_1^* + r_1 K_{P,1} = 0 \tag{B.2.0.6}$$

Solving for S_1^*, we obtain the expression:

$$S_1^* = \frac{(1 + K_{K,1} + r_1 K_{P,1}) - \sqrt{(1 + K_{K,1} + r_1 K_{P,1})^2 - 4 r_1 K_{P,1}}}{2} \tag{B.2.0.7}$$

where the positive branch of the quadratic solution is chosen to obtain physically realistic values of S_1^* (i.e. $0 \leq S_1^* \leq 1$). We can show that S_1^* for a 2K1P with an unsaturated phosphatase is strictly hyperbolic in r_1 by taking the second derivative of B.2.0.7 with respect to r_1:

$$\frac{d^2 S_1^*}{d r_1^2} = -\frac{2 K_{K,1} K_{P,1}^2}{(-4 r_1 K_{P,1} + (1 + K_{K,1} + r_1 K_{P,1})^2)^{3/2}} \tag{B.2.0.8}$$

Note that, for positive real values of both the rate constants and the concentrations, both the numerator and denominator in the above equation are positive. The second derivative is thus always negative (i.e. the curvature is concave), and the variation of S_1^* with r_1 lacks an inflection point. As a result, a system with unsaturated phosphatases cannot exhibit the sigmoidal behavior characteristic of an ultrasensitive, switch-like response.

B.3 Analytical solution for 1-Kinase/1-Substrate Model with Synthesis and Degradation

From the ODEs derived in B.1.4, we have at steady state:

$$\frac{d[K]}{dt} = [KS]k_- + [KS]k_{cat} + [KS]k_{deg,U} - [K][S]k_+ = 0 \tag{B.3.0.1}$$

$$\frac{d[S]}{dt} = [KS]k_- + k_{synth} - [K][S]k_+ - [S]k_{deg,U} = 0 \tag{B.3.0.2}$$

$$\frac{d[S^*]}{dt} = [KS]k_{cat} - [S^*]k_{deg,P} = 0 \tag{B.3.0.3}$$

$$\frac{d[KS]}{dt} = [K][S]k_+ - [KS]k_- - [KS]k_{cat} - [KS]k_{deg,U} = 0 \tag{B.3.0.4}$$

Additionally, we can define the conservation of mass of the kinase K:

$$[K]_0 = [K] + [KS] \tag{B.3.0.5}$$

From B.3.0.4 we can define the concentration of KS as:

$$[KS] = \frac{[K][S]k_+}{k_- + k_{cat} + k_{deg,U}} \tag{B.3.0.6}$$

Equation B.3.0.6 can then be substituted into B.3.0.5 to define the concentration of free kinase K:

$$\begin{aligned}
[K]_0 &= [K] + \frac{[K][S]k_+}{k_- + k_{cat} + k_{deg,U}} \\
[K]_0 &= [K]\left(1 + \frac{[S]k_+}{k_- + k_{cat} + k_{deg,U}}\right) \\
[K] &= \frac{[K]_0}{1 + \frac{[S]k_+}{k_- + k_{cat} + k_{deg,U}}}
\end{aligned} \tag{B.3.0.7}$$

We can then substitute B.3.0.7 into B.3.0.6 to get:

$$[KS] = \frac{[K]_0[S]}{\frac{k_- + k_{cat} + k_{deg,U}}{k_+} + [S]} \tag{B.3.0.8}$$

We can simplify B.3.0.8 by defining:

$$K_{M,deg} \equiv \frac{k_- + k_{cat} + k_{deg,U}}{k_+}$$

where $K_{M,deg}$ is the analogue of the Michaelis-Menten constant for the kinase K, taking into account the effects of substrate degradation. Equation B.3.0.8 then can be written:

$$[KS] = \frac{[K]_0[S]}{K_{M,deg} + [S]} \tag{B.3.0.9}$$

From B.3.0.3 we obtain the expression:

$$[KS]k_{cat} = [S^*]k_{deg,P}$$

into which we can substitute B.3.0.9 to get:

$$\frac{[K]_0[S]k_{cat}}{K_{M,deg} + [S]} = [S^*]k_{deg,P} \tag{B.3.0.10}$$

We can then multiply both sides of B.3.0.10 by $[S]_0/[S]_0$:

$$\frac{[K]_0 S \cdot k_{cat}}{K_{deg} + S} = S^* k_{deg,P}[S_0]$$

where $S \equiv [S]/[S]_0$, $S^* \equiv [S^*]/[S]_0$, and $K_{deg} \equiv K_{M,deg}/[S]_0$. We can then define $r_{deg} \equiv [K]_0 k_{cat}/[S]_0 k_{deg,P}$, the ratio of the maximum velocity of the kinase to the maximum velocity of phosphorylated sub-

strate degradation, to get:

$$\frac{r_{deg}S}{K_{deg}+S} = S^* \tag{B.3.0.11}$$

Following the standard Michaelis-Menten assumptions [26, 72], we have $[S]_0 \gg [K]_0$. This gives us $[S]_0 = [S] + [S^*]$, or $1 = S + S^*$, which can be substituted into B.3.0.11:

$$\begin{aligned} \frac{r_{deg}(1-S^*)}{K_{deg}+1-S^*} &= S^* \\ r_{deg} - r_{deg}S^* &= K_{deg}S^* + S^* - (S^*)^2 \\ (S^*)^2 - (1+r_{deg}+K_{deg})S^* + r_{deg} &= 0 \end{aligned} \tag{B.3.0.12}$$

We can then solve B.3.0.12 for S^*:

$$S^* = \frac{1 + r_{deg} + K_{deg} - \sqrt{(1+r_{deg}+K_{deg})^2 - 4r_{deg}}}{2} \tag{B.3.0.13}$$

where we have again chosen the positive branch of the solution to ensure $0 \leq S_1^* \leq 1$. From this derivation, we can see that degradation has two effects on the fraction of phosphorylated substrate. The degradation rate of unphosphorylated substrate can change the saturation of the kinase K by the substrate through altering the Michaelis-Menten-like constant $K_{M,deg}$, in the same way as altering the dissociation or catalytic rates. Additionally, the degradation rate of phosphorylated substrate can modify the magnitude of the fraction of phosphorylated substrate by changing r_{deg}.

We can show that S^* is strictly hyperbolic in r_{deg} by taking the second derivative of B.3.0.13 with regard to r_{deg}:

$$\frac{d^2S^*}{dr_{deg}^2} = -\frac{2K_{deg}}{(-4r_{deg} + (1+r_{deg}+K_{deg})^2)^{3/2}} \tag{B.3.0.14}$$

As for the response with an unsaturated phosphatase (equation B.2.0.8), both the numerator and denominator of the above equation are strictly positive, so the second derivative is always negative.

A system that relies on degradation to achieve effective dephosphorylation thus has no inflection point in r_{deg} and cannot exhibit a sigmoidal response to signals.

B.4 UniProt Data

We searched the UniProt database for the number of serine/threonine and tyrosine kinases and phosphatases found in all complete eukaryotic genomes [89]. For each genome, we searched for UniProt for reviewed entries that included the enzyme classification numbers for kinases and phosphatases (see Table B.1). We then counted the number of entries for each genome in the search results. In order to prevent genomes with small numbers of reviewed kinases or phosphatases from unduly influencing our results, we ignored genomes with less than 5 phosphatases or 5 kinases for any given residue class (i.e. serine/threonine or tyrosine). This resulted in 16 genomes for serine/threonine enzymes (See Table B.2) and 9 genomes for tyrosine enzymes (See Table B.3).

Additionally, we used UniProt to determine the number of phosphoproteins in complete eukaryotic genomes. We searched UniProt for reviewed entries with keyword 'Phosphoprotein'. We then analyzed the search results and computed the number of entries for each of the 16 species from the serine/threonine enzyme results and the 9 species from the tyrosine enzyme results. The phosphatase numbers represent the total number of phosphatases from tables B.2 and B.3.

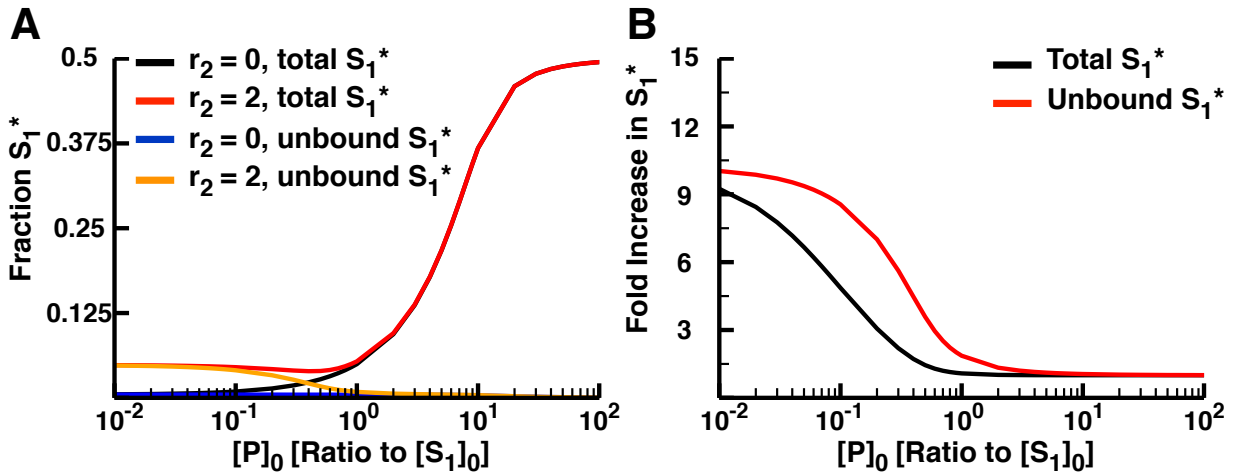

Figure B.1: The effects of increased phosphatase concentration on substrate crosstalk with strong $K_{M,P,i}$s (*A*) The fraction S_1^* as a function of the concentration of the phosphatase. The concentration of the kinase is increased in order to maintain the values of r_1 and r_2 with constant catalytic rates for different concentrations of the phosphatase. At low concentrations of P, S_1 phosphorylation is increased by activation of the S_2 pathway, moving from $r_2 = 0$ (black) to $r_2 = 2$ (red). As P is expressed in concentrations greater than the substrates, the difference between the curves is removed. However, the fraction S_1^* is greatly increased. Additionally, the fraction of unbound S_1^* decreases with $[P]_0$, indicating that the increase in total fraction S_1^* is likely due to it being bound to the phosphatase. (*B*) The fold increase in S_1^* as a function of the concentration of the phosphatase. As stated above, the crosstalk between S_1 and S_2 is removed when the phosphatase is present in concentrations larger than those of the substrates.

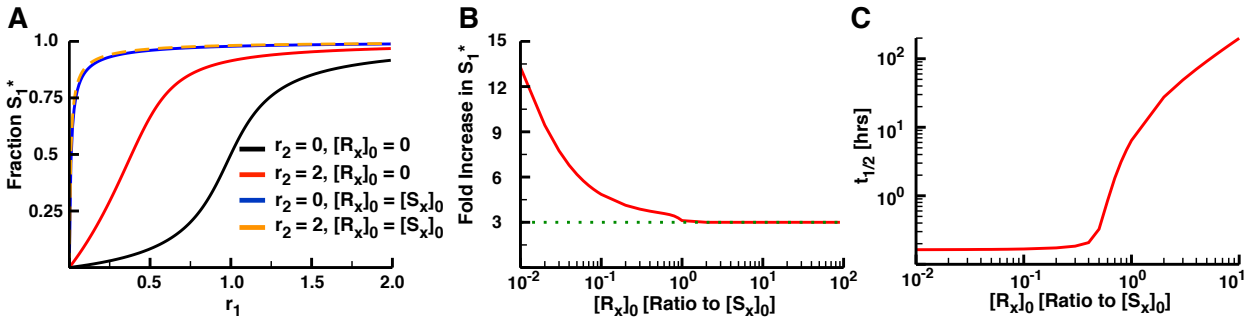

Figure B.2: The effects of reservoir proteins on substrate crosstalk (*A*) The fraction S_1^* as a function of r_1. Without the reservoir proteins, S_1 responds to signals from S_2 (red versus black curves). The crosstalk is removed with the addition of the reservoir proteins (orange versus blue curves). Note, however, that the response becomes hyperbolic in r_1. (*B*) The fold increase in S_1^* as a function of reservoir protein concentration. As the concentrations of the reservoir proteins increases, the crosstalk between the substrates is gradually removed. (*C*) The half-life of S_1 phosphorylation as a function of the concentration of reservoir proteins. Note that when the reservoir proteins are at stoichiometric or greater concentrations, the time required to completely dephosphorylate the substrates greatly increases.

Table B.1: Enzyme classification numbers used to search UniProt

Enzyme	E.C. Number
Serine/Threonine Phosphatases	3.1.3.3 / 3.1.3.16
Serine/Threonine Kinases	2.7.11.x
Tyrosine Phosphatases	3.1.3.48
Tyrosine Kinases	2.7.10.x

Table B.2: The numbers and ratios of serine/threonine kinases and phosphatases from UniProt used in Figure 3.1 of Chapter 3.

Species	Serine/Threonine Phosphatases	Serine/Threonine Kinases	Ratio
Arabidopsis thaliana	115	559	0.206
Bos taurus	26	81	0.321
Caenohabditis elegans	15	89	0.169
Danio rerio	14	40	0.350
Dictyostelium discoideum	21	222	0.095
Drosophila melanogaster	22	66	0.333
Gallus gallus	9	36	0.250
Homo sapiens	79	372	0.212
Mus musculus	72	374	0.193
Oryctolagus cuniculus	7	23	0.304
Oryza sativa	94	120	0.783
Pongo abelii	10	37	0.270
Rattus norvegivus	40	188	0.213
Saccharomyces cerevisiae	24	124	0.194
Schizosaccharomyces pombe	20	100	0.200
Xenopus laevis	17	72	0.236

Table B.3: The numbers and ratios of tyrosine kinases and phosphatases from UniProt used in Figure 3.1 of Chapter 3.

Species	Tyrosine Phosphatases	Tyrosine Kinases	Ratio
Arabidopsis thaliana	9	16	0.563
Bos taurus	13	9	1.444
Caenohabditis elegans	12	11	1.091
Drosophila melanogaster	14	21	0.667
Gallus gallus	11	32	0.344
Homo sapiens	87	95	0.916
Mus musculus	81	95	0.853
Rattus norvegivus	39	49	0.796
Xenopus laevis	11	25	0.440

Table B.4: The numbers and ratios of the total number of phosphatases and substrates from UniProt used in Figure 3.1 of Chapter 3.

Species	Total Phosphatases	Total Substrates	Ratio
Arabidopsis thaliana	124	1295	10.444
Bos taurus	39	1759	45.103
Caenohabditis elegans	27	89	3.296
Danio rerio	14	195	13.929
Dictyostelium discoideum	21	160	7.619
Drosophila melanogaster	36	833	23.139
Gallus gallus	20	282	14.100
Homo sapiens	166	5924	35.687
Mus musculus	153	5313	34.725
Oryctolagus cuniculus	7	282	40.286
Oryza sativa	94	150	1.596
Pongo abelii	10	819	81.900
Rattus norvegivus	79	2691	34.063
Saccharomyces cerevisiae	24	2425	101.042
Schizosaccharomyces pombe	20	1067	53.350
Xenopus laevis	28	277	9.893

Appendix C

Appendix for Crosstalk and the Evolution of Specificity in Two-Component Signaling

C.1 Analytical Results for Competition in TCS

C.1.1 Steady-state solution for RR_1^*

The first model we considered was aimed at characterizing RR_1 phosphorylation when a single HK acts as both a phosphotransferase and a phosphatase for multiple substrates RR_i, $i = 1,...,N$. Our model for this system includes the following reactions:

$$HK \underset{k_b}{\overset{k_f}{\rightleftharpoons}} HK^*$$

$$HK^* + RR_i \underset{k_{-,1,i}}{\overset{k_{+,1,i}}{\rightleftharpoons}} HK^*RR_i \xrightarrow{k_{cat,1,i}} HK + RR_i^*$$

$$HK + RR_i^* \underset{k_{-,2,i}}{\overset{k_{+,2,i}}{\rightleftharpoons}} HKRR_i^* \xrightarrow{k_{cat,2,i}} HK + RR_i$$

where the parameters can, in principle, vary independently for all N substrates. This model is taken directly from Batchelor and Goulian [32]. We also considered the more complex model from Ray, *et al.* [33] in which the phosphotransfer reaction produces an $HKRR_i^*$ complex that can

then either dissociate or undergo the phosphatase catalytic step. We found that the two models behaved similarly in the presence of competing *RR*s for the parameters defined in Ray, *et al.* [33]. We thus focused on the simpler model due to its analytical tractability. Note that these models do not include a $HK^*RR_i^*$ complex due to electrostatic clashes between the two phosphate groups. Based on the chemical reactions above, we obtain the following kinetic equations:

$$\frac{d[HK]}{dt} = [HK^*]k_b + \sum_i [HK^*RR_i]k_{cat,1,i} + \sum_i ([HKRR_i^*](k_{-,2,i} + k_{cat,2,i})) - [HK]k_f - \sum_i [HK][RR_i^*]k_{+,2,i} \quad \text{(C.1.1.1)}$$

$$\frac{d[HK^*]}{dt} = [HK]k_f + \sum_i [HK^*RR_i]k_{-,1,i} - [HK^*]k_b - \sum_i [HK^*][RR_i]k_{+,1,i} \quad \text{(C.1.1.2)}$$

$$\frac{d[RR_i]}{dt} = [HK^*RR_i]k_{-,1,i} + [HKRR_i^*]k_{cat,2,i} - [HK^*][RR_i]k_{+,1,i} \quad \text{(C.1.1.3)}$$

$$\frac{d[HK^*RR_i]}{dt} = [HK^*][RR_i]k_{+,1,i} - [HK^*RR_i](k_{-,1,i} + k_{cat,1,i}) \quad \text{(C.1.1.4)}$$

$$\frac{d[RR_i^*]}{dt} = [HK^*RR_i]k_{cat,1,i} + [HKRR_i^*]k_{-,2,i} - [HK][RR_i^*]k_{+,2,i} \quad \text{(C.1.1.5)}$$

$$\frac{d[HKRR_i^*]}{dt} = [HK][RR_i^*]k_{+,2,i} - [HKRR_i^*](k_{-,2,i} + k_{cat,2,i}) \quad \text{(C.1.1.6)}$$

From C.1.1.4 and C.1.1.6 we can define:

$$[HK^*RR_i] = \frac{[HK^*][RR_i]}{K_{m,1,i}} \quad \text{(C.1.1.7)}$$

$$[HKRR_i^*] = \frac{[HK][RR_i^*]}{K_{m,2,i}} \quad \text{(C.1.1.8)}$$

where $K_{m,j,i} = (k_{-,j,i} + k_{cat,j,i})/k_{+,j,i}$, $i = 1 \ldots N$, j = 1 or 2. Substituting C.1.1.7 into C.1.1.3 we get at steady-state:

$$\frac{d[RR_i]}{dt} = [HK^*RR_i]k_{-,1,i} + [HKRR_i^*]k_{cat,2,i} - [HK^*RR_i]k_{+,1,i}K_{m,1,i} = 0$$

$$[HK^*RR_i]k_{cat,1,i} = [HKRR_i^*]k_{cat,2,i} \quad \text{(C.1.1.9)}$$

Additionally, by substituting C.1.1.7 into C.1.1.2 we can obtain an expression for $[HK]$ at steady state:

$$\begin{aligned}
\frac{d[HK^*]}{dt} &= [HK]k_f + \sum_i [HK^*][RR_i]\frac{k_{-,1,i}}{K_{m,1,i}} - [HK^*]k_b - \sum_i [HK^*][RR_i]k_{+,1,i} = 0 \\
&[HK]k_f + [HK^*]\left(\sum_i [RR_i]\left(\frac{k_{-,1,i}}{K_{m,1,i}} - k_{+,1,i}\right) - k_b\right) = 0 \\
&[HK]k_f - [HK^*]\left(\sum_i [RR_i]\frac{k_{cat,1,i}}{K_{m,1,i}} + k_b\right) = 0 \\
&[HK] = [HK^*]\left(\frac{\sum_i [RR_i]\frac{k_{cat,1,i}}{K_{m,1,i}} + k_b}{k_f}\right)
\end{aligned} \tag{C.1.1.10}$$

Similarly we can substitute C.1.1.8 and C.1.1.9 into C.1.1.1 to obtain an expression for $[HK^*]$ at steady state:

$$\begin{aligned}
\frac{d[HK]}{dt} &= [HK^*]k_b + \sum_i [HK][RR_i^*]\frac{k_{cat,2,i}}{K_{m,2,i}} + \sum_i [HK][RR_i^*]\frac{(k_{-,2,i} + k_{cat,2,i})}{K_{m,2,i}} - [HK]k_f \\
&\quad - \sum_i [HK][RR_i^*]k_{+,2,i} = 0 \\
&[HK^*]k_b + [HK]\left(\sum_i [RR_i^*]\frac{k_{cat,2,i}}{K_{m,2,i}} - k_f\right) = 0 \\
&[HK^*] = [HK]\left(\frac{-\sum_i [RR_i^*]\frac{k_{cat,2,i}}{K_{m,2,i}} + k_f}{k_b}\right)
\end{aligned} \tag{C.1.1.11}$$

The conservation of mass for the histidine kinase is:

$$[HK]_0 = [HK] + [HK^*] + \sum_i [HKRR_i^*] + \sum_i [HK^*RR_i] \tag{C.1.1.12}$$

By substituting C.1.1.8, C.1.1.9 and C.1.1.11 into C.1.1.12 we obtain:

$$[HK]_0 = [HK] + [HK]\left(\frac{-\sum_i [RR_i^*]\left(\frac{k_{cat,2,i}}{K_{m,2,i}}\right) + k_f}{k_b}\right) + \sum_i [HK]\frac{[RR_i^*]}{K_{m,2,i}} + \sum_i [HK]\frac{[RR_i^*]}{K_{m,2,i}} \cdot \frac{k_{cat,2,i}}{k_{cat,1,i}}$$

$$[HK]_0 = [HK]\left(1 + \frac{-\sum_i [RR_i^*]\left(\frac{k_{cat,2,i}}{K_{m,2,i}}\right) + k_f}{k_b} + \sum_i \frac{[RR_i^*]}{K_{m,2,i}}\left(1 + \frac{k_{cat,2,i}}{k_{cat,1,i}}\right)\right)$$

$$[HK]_0 = [HK]\left(1 + \frac{k_f}{k_b} + \sum_i \frac{[RR_i^*]}{K_{m,2,i}}\left(1 + \frac{k_{cat,2,i}}{k_{cat,1,i}} - \frac{k_{cat,2,i}}{k_b}\right)\right)$$

which can be simplified by defining:

$$\alpha_1 \equiv 1 + \frac{k_f}{k_b}, \quad \beta_i \equiv 1 + \frac{k_{cat,2,i}}{k_{cat,1,i}} - \frac{k_{cat,2,i}}{k_b}$$

where $i = 1, \ldots, N$. This gives us:

$$[HK]_0 = [HK]\left(\alpha_1 + \sum_i \beta_i \frac{[RR_i^*]}{K_{m,2,i}}\right)$$

$$[HK] = \frac{[HK]_0}{\alpha_1 + \sum_i \beta_i \frac{[RR_i^*]}{K_{m,2,i}}}$$

$$[HKRR_1^*] = \frac{[HK]_0[RR_1^*]}{\left(\alpha_1 + \sum_{j=2}^{N} \beta_j \frac{[RR_j^*]}{K_{m,2,j}}\right) K_{m,2,1} + \beta_1 [RR_1^*]}$$

which can be further simplified by defining:

$$\varepsilon_1 \equiv \alpha_1 + \sum_{j=2}^{N} \beta_j \frac{[RR_j^*]}{K_{m,2,j}}$$

finally producing:

$$[HKRR_1^*] = \frac{[HK]_0[RR_1^*]}{\varepsilon_1 K_{m,2,1} + \beta_1 [RR_1^*]} \tag{C.1.1.13}$$

Similarly we can substitute C.1.1.7,C.1.1.8 and C.1.1.10 into C.1.1.12 to obtain:

$$[HK]_0 = [HK^*]\left(\frac{\sum_i [RR_i]\frac{k_{cat,1,i}}{K_{m,1,i}} + k_b}{k_f}\right) + [HK^*] + \sum_i [HK^*]\frac{[RR_i]}{K_{m,1,i}}\left(\frac{k_{cat,1,i}}{k_{cat,2,i}}\right) + \sum_i [HK^*]\frac{[RR_i]}{K_{m,1,i}}$$

$$[HK]_0 = [HK^*]\left(1 + \frac{\sum_i [RR_i]\frac{k_{cat,1,i}}{K_{m,1,i}} + k_b}{k_f} + \sum_i \frac{[RR_i]}{K_{m,1,i}}\left(1 + \frac{k_{cat,1,i}}{k_{cat,2,i}}\right)\right)$$

$$[HK]_0 = [HK^*]\left(1 + \frac{k_b}{k_f} + \sum_i \frac{[RR_i]}{K_{m,1,i}}\left(1 + \frac{k_{cat,1,i}}{k_{cat,2,i}} + \frac{k_{cat,1,i}}{k_f}\right)\right)$$

which can be simplified by defining:

$$\alpha_1' \equiv 1 + \frac{k_b}{k_f}, \quad \beta_i' \equiv 1 + \frac{k_{cat,1,i}}{k_{cat,2,i}} + \frac{k_{cat,1,i}}{k_f}$$

where $i = 1, \ldots, N$. This gives us:

$$[HK]_0 = [HK^*]\left(\alpha_1' + \sum_i \beta_i' \frac{[RR_i]}{K_{m,1,i}}\right)$$

$$[HK^*] = \frac{[HK]_0}{\alpha_1' + \sum_i \beta_i' \frac{[RR_i]}{K_{m,1,i}}}$$

$$[HK^*RR_1] = \frac{[HK]_0[RR_1]}{\left(\alpha_1' + \sum_{j=2}^{N} \beta_j' \frac{[RR_j]}{K_{m,1,j}}\right) K_{m,1,1} + \beta_1'[RR_1]}$$

which can be further simplified by defining:

$$\varepsilon_1' \equiv \alpha_1' + \sum_{j=2}^{N} \beta_j' \frac{[RR_j]}{K_{m,1,j}}$$

finally giving us:

$$[HK^*RR_1] = \frac{[HK]_0[RR_1]}{\varepsilon_1' K_{m,1,1} + \beta_1'[RR_1]} \tag{C.1.1.14}$$

We can then substitute C.1.1.13 and C.1.1.14 into C.1.1.9 to get:

$$\frac{k_{cat,1,1}[HK]_0[RR_1]}{\varepsilon'_1 K_{m,1,1}+\beta'_1[RR_1]} = \frac{k_{cat,2,1}[HK]_0[RR^*_1]}{\varepsilon_1 K_{m,2,1}+\beta_1[RR^*_1]}$$
$$\frac{r_1 RR_1}{\varepsilon'_1 K_{1,1}+\beta'_1 RR_1} = \frac{RR^*_1}{\varepsilon_1 K_{2,1}+\beta_1 RR^*_1} \tag{C.1.1.15}$$

where $r_1 \equiv \frac{k_{cat,1,1}[HK]_0}{k_{cat,2,1}[HK]_0} = \frac{V_{max,1,1}}{V_{max,2,1}} = \frac{k_{cat,1,1}}{k_{cat,2,1}}$, $RR_1 \equiv \frac{[RR_1]}{[RR_1]_0}$, $RR^*_1 = \frac{[RR^*_1]}{[RR_1]_0}$, $K_{1,1} \equiv \frac{K_{m,1,1}}{[RR_1]_0}$, and $K_{2,1} \equiv \frac{K_{m,2,1}}{[RR_1]_0}$. Note that C.1.1.5 is similar to the Goldbeter-Koshland solution [26]. Next, we make the standard Michaelis assumption $[RR_1]_0 \approx [RR_1] + [RR^*_1]$, which we can apply to C.1.1.15:

$$\frac{r_1(1-RR^*_1)}{\varepsilon'_1 K_{1,1}+\beta'_1(1-RR^*_1)} = \frac{RR^*_1}{\varepsilon_1 K_{2,1}+\beta_1 RR^*_1}$$
$$\varepsilon'_1 K_{1,1} RR^*_1 + \beta'_1 RR^*_1 - \beta'_1 (RR^*_1)^2 = \varepsilon_1 r_1 K_{2,1} - \varepsilon_1 r_1 K_{2,1} RR^*_1 + r_1 \beta_1 RR^*_1 - r_1 \beta_1 (RR^*_1)^2$$
$$(r_1\beta_1 - \beta'_1)(RR^*_1)^2 - (r_1\beta_1 - \beta'_1 - \varepsilon'_1 K_{1,1} - \varepsilon_1 r_1 K_{2,1}) RR^*_1 - \varepsilon_1 r_1 K_{2,1} = 0 \tag{C.1.1.16}$$

We can then solve C.1.1.16 for RR^*_1:

$$RR^*_1 = \frac{(r_1\beta_1 - \beta'_1) - (\varepsilon'_1 K_{1,1} + \varepsilon_1 r_1 K_{2,1}) \pm \sqrt{((r_1\beta_1 - \beta'_1) - (\varepsilon'_1 K_{1,1} + \varepsilon_1 r_1 K_{2,1}))^2 + 4(r_1\beta_1 - \beta'_1)\varepsilon_1 r_1 K_{2,1}}}{2(r_1\beta_1 - \beta'_1)} \tag{C.1.1.17}$$

Note that when RR_1 has no competition (i.e. $N = 1$), $\varepsilon_1 = \alpha_1$ and $\varepsilon'_1 = \alpha'_1$.

C.1.2 $r_1\beta_1 - \beta_1 < 0$

We can show that the denominator of C.1.1.17 is negative, i.e. $r_1\beta_1 - \beta_1 < 0$. First we will assume the opposite:

$$r_1\beta_1 - \beta'_1 \geq 0$$

By substituting the full expressions for each term we then get:

$$\begin{aligned}\frac{k_{cat,1,1}}{k_{cat,2,1}}\left(1+\frac{k_{cat,2,1}}{k_{cat,1,1}}-\frac{k_{cat,2,1}}{k_b}\right)-1-\frac{k_{cat,1,1}}{k_{cat,2,1}}-\frac{k_{cat,1,1}}{k_f}&\geq 0\\ \frac{k_{cat,1,1}}{k_{cat,2,1}}+1-\frac{k_{cat,1,1}}{k_b}-1-\frac{k_{cat,1,1}}{k_{cat,2,1}}-\frac{k_{cat,1,1}}{k_f}&>0\\ -\frac{k_{cat,1,1}}{k_b}-\frac{k_{cat,1,1}}{k_f}&\geq 0\end{aligned} \tag{C.1.2.1}$$

Since all the rate parameters are positive real numbers, it is clear that C.1.2.1 is impossible, indicating that $r_1\beta_1-\beta_1'<0$.

C.1.3 Only the positive branch of RR_1^* is realistic

Since RR_1^* represents a molar fraction of modified substrate, we know that $0\leq RR_1^*\leq 1$. The solution for RR_1^* (C.1.1.17) has two branches, a positive and negative branch. We can demonstrate that the negative branch will violate the previously stated bounds, restricting the solution for RR_1^* to the positive branch for a system with no competing response regulators ($N=1$). To make the derivation easier to follow we first define:

$$x=-((r_1\beta_1-\beta_1')-(\alpha_1'K_{1,1}+\alpha_1 r_1 K_{2,1}))>0,\quad y=4(r_1\beta_1-\beta_1')\alpha_1 r_1 K_{2,1}<0,\quad z=r_1\beta_1-\beta_1<0$$

which simplifies C.1.1.17 to:

$$RR_1^*=\frac{-x\pm\sqrt{x^2+y}}{2z} \tag{C.1.3.1}$$

Now, we can show the negative branch of RR_1^* violates its upper bounds by assuming the opposite:

$$\begin{aligned}RR_1^*=\frac{-x-\sqrt{x^2+y}}{2z}&\leq 1\\ -x-\sqrt{x^2+y}&\geq 2z\\ \sqrt{x^2+y}&\leq -x-2z\end{aligned} \tag{C.1.3.2}$$

If the right hand side of C.1.3.2 is negative then we have already reached a contradiction since the left hand side must be positive and real. If the right hand side is positive, then we can square both sides without loss of information:

$$x^2 + y \leq x^2 + 4xz + 4z^2$$
$$y \leq 4xz + 4z^2$$

By their definitions, x and y can be expressed as $x = -z + \alpha_1' K_{1,1} + \alpha_1 r_1 K_{2,1}$ and $y = 4z\alpha_1 r_1 K_{2,1}$. This gives us:

$$\begin{aligned} 4z\alpha_1 r_1 K_{2,1} &\leq 4z(-z + \alpha_1' K_{1,1} + \alpha_1 r_1 K_{2,1}) + 4z^2 \\ \alpha_1 r_1 K_{2,1} &\geq -z + \alpha_1' K_{1,1} + \alpha_1 r_1 K_{2,1} + z \\ 0 &\geq \alpha_1' K_{1,1} \end{aligned} \tag{C.1.3.3}$$

Which is clearly impossible, allowing the conclusion that the negative branch of RR_1^* violates the upper bounds. Similarly we can show that the positive branch of RR_1^* does not violate its bounds. First we can test the upper bound, and show that $RR_1^* \neq 1$ by assuming the opposite:

$$\begin{aligned} RR_1^* = \frac{-x + \sqrt{x^2 + y}}{2z} &\geq 1 \\ -x + \sqrt{x^2 + y} &\leq 2z \\ \sqrt{x^2 + y} &\leq x + 2z \end{aligned} \tag{C.1.3.4}$$

If the right hand side of C.1.3.4 is negative then we have already reach a contradiction. Otherwise we can square both sides without loss of information:

$$x^2 + y \leq x^2 + 4xz + 4z^2 \tag{C.1.3.5}$$

which we have previously shown leads to the contradiction shown in C.1.3.3, confirming that the positive branch of RR_1^* does not violate the upper bound and $RR_1^* < 1$. Next we can demonstrate that the positive branch of RR_1^* does not violate its lower bounds by assuming:

$$\begin{aligned} RR_1^* = \frac{-x+\sqrt{x^2+y}}{2z} &< 0 \\ -x+\sqrt{x^2+y} &> 0 \\ \sqrt{x^2+y} &> x \end{aligned} \tag{C.1.3.6}$$

By the definition of x we know that the right hand side of C.1.3.6 has to be positive, so we can square both sides of C.1.3.6 without loss of information:

$$\begin{aligned} x^2+y &> x^2 \\ y &> 0 \end{aligned} \tag{C.1.3.7}$$

We know by definition $y < 0$, showing that C.1.3.7 to be impossible. This confirms that RR_1^* does not violate its lower and upper bounds. We can then simplify C.1.1.17 for $N = 1$ to:

$$\begin{aligned} RR_1^* &= \frac{-x+\sqrt{x^2+y}}{2z} \\ &= \frac{(r_1\beta_1-\beta_1') - (\alpha_1' K_{1,1} + \alpha_1 r_1 K_{2,1}) + \sqrt{((r_1\beta_1-\beta_1') - (\alpha_1' K_{1,1} + \alpha_1 r_1 K_{2,1}))^2 + 4(r_1\beta_1-\beta_1')\alpha_1 r_1 K_{2,1}}}{2(r_1\beta_1-\beta_1')} \end{aligned} \tag{C.1.3.8}$$

We constructed a version of this model in CVODE order to compare C.1.3.8 with numerical simulations of the underlying ODEs (C.1.1.1-C.1.1.6). Using the relevant parameters described in section C.3, we considered the model for a single HK and RR with $[RR]_0$ at three different concentrations and simulated each system to steady state. Comparison of the RR^* vs. k_f/k_b curves from these simulations and C.1.3.8 showed excellent agreement (Fig. C.1).

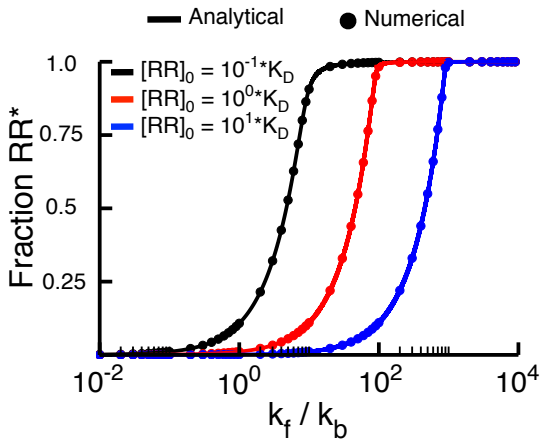

Figure C.1: Fraction RR^* versus k_f/k_b comparing the analytical solution C.1.3.8 and numerical simulations run to steady state. In both cases, the relevant kinetic parameters were taken from the tables in section C.3.

C.2 The effects of increased *RR* concentration on TCS

C.2.1 $d[RR_1^*]/d[RR_1]_0 > 0$

We can show that increasing the concentration of RR_1 causes a smaller fraction of RR_1 to be phosphorylated while still increasing the concentration of RR_1^*. Here we assume that there are no competing substrates, so $\varepsilon_1 = \alpha_1$, $\varepsilon_1' = \alpha_1'$, and $\frac{d\alpha_1}{d[RR_1]_0} = \frac{d\alpha_1'}{d[RR_1]_0} = 0$. We get from C.1.1.15:

$$\frac{r_1[RR]}{\alpha_1' K_{m,1,1} + \beta_1'[RR_1]} = \frac{[RR_1^*]}{\alpha_1 K_{m,2,1} + \beta_1[RR_1^*]} \tag{C.2.1.1}$$

We make the Michaelis-Menten assumption $[RR_1]_0 \approx [RR_1] + [RR_1^*]$, which we can substitute into C.2.1.1 to get:

$$\frac{r_1([RR_1]_0 - [RR^*])}{\alpha_1' K_{m,1,1} + \beta_1'([RR_1]_0 - [RR_1^*])} = \frac{[RR_1^*]}{\alpha_1 K_{m,2,1} + \beta_1[RR_1^*]} \tag{C.2.1.2}$$

We can take the implicit derivative of C.2.1.2 with respect to $[RR_1]_0$ to obtain:

$$\frac{-\alpha_1' r_1 K_{m,1,1}\left(\frac{d[RR_1^*]}{d[RR_1]_0}-1\right)}{(\alpha_1' K_{m,1,1}+\beta_1'([RR_1]_0-[RR_1^*]))^2} = \frac{\alpha_1 K_{m,2,1}\frac{d[RR_1^*]}{d[RR_1]_0}}{(\alpha_1 K_{m,2,1}+\beta_1[RR_1^*])^2}$$

$$\frac{\alpha_1' r_1 K_{m,1,1}}{(\alpha_1' K_{m1,1}+\beta_1'([RR_1]_0-[RR_1^*]))^2} = \left(\frac{\alpha_1' r_1 K_{m,1,1}}{(\alpha_1' K_{m1,1}+\beta_1'([RR_1]_0-[RR_1^*]))^2}+\frac{\alpha_1 K_{m,2,1}}{(\alpha_1 K_{m,2,1}+\beta_1[RR_1^*])^2}\right)\frac{d[RR_1^*]}{d[RR_1]_0}$$

Which we can solve for $\frac{d[RR_1^*]}{d[RR_1]_0}$:

$$\frac{d[RR_1^*]}{d[RR_1]_0} = \frac{\frac{\alpha_1' r_1 K_{m,1,1}}{(\alpha_1' K_{m1,1}+\beta_1'([RR_1]_0-[RR_1^*]))^2}}{\frac{\alpha_1' r_1 K_{m,1,1}}{(\alpha_1' K_{m1,1}+\beta_1'([RR_1]_0-[RR_1^*]))^2}+\frac{\alpha_1 K_{m,2,1}}{(\alpha_1 K_{m,2,1}+\beta_1[RR_1^*])^2}} \tag{C.2.1.3}$$

Since by their definitions $\alpha_1 > 0$, $\alpha_1' > 0$, $r_1 > 0$, $K_{m,x,1} > 0$, and since $x^2 > 0 \; \forall \; x \in \mathbb{R}$, the right hand side of C.2.1.3 positive, and thus $\frac{d[RR_1^*]}{d[RR_1]_0} > 0$.

C.2.2 $dRR_1^*/d[RR_1]_0 < 0$

Next we considered the behavior of $\frac{dRR_1^*}{d[RR_1]_0}$. To do this we can take the derivative of C.1.3.8, once again without competing *RR*s, with respect to $[RR_1]_0$:

$$\frac{dRR_1^*}{d[RR_1]_0} = \frac{\alpha_1' K_{1,1}+\alpha_1 r_1 K_{2,1}+\frac{(\alpha_1' K_{1,1}+\alpha_1 r_1 K_{2,1})((r_1\beta_1-\beta_1')-(\alpha_1' K_{1,1}+\alpha_1 r_1 K_{2,1}))-2(r_1\beta_1-\beta_1')\alpha_1 r_1 K_{2,1}}{\sqrt{((r_1\beta_1-\beta_1')-(\alpha_1' K_{1,1}+\alpha_1 r_1 K_{2,1}))^2+4(r_1\beta_1-\beta_1')\alpha_1 r_1 K_{2,1}}}}{2(r_1\beta_1-\beta_1')[RR_1]_0} \tag{C.2.2.1}$$

Using the previously defined x, y and z we can simplify C.2.2.1 to:

$$\frac{dRR_1^*}{d[RR_1]_0} = \frac{\alpha_1' K_{1,1}+\alpha_1 r_1 K_{2,1}+\frac{(\alpha_1' K_{1,1}+\alpha_1 r_1 K_{2,1})(-x)-2z\alpha_1 r_1 K_{2,1}}{\sqrt{x^2+y}}}{2z[RR_1]_0} \tag{C.2.2.2}$$

We can then bring the $\sqrt{x^2+y}$ term to the denominator to get:

$$\frac{dRR_1^*}{d[RR_1]_0} = \frac{\alpha_1' K_{1,1} + \alpha_1 r_1 K_{2,1}}{[RR_1]_0\sqrt{x^2+y}}\left(\frac{-x+\sqrt{x^2+y}}{2z}\right) - \frac{\alpha_1 r_1 K_{2,1}}{[RR_1]_0\sqrt{x^2+y}}$$
$$\frac{dRR_1^*}{d[RR_1]_0} = \frac{\alpha_1' K_{1,1} + \alpha_1 r_1 K_{2,1}}{[RR_1]_0\sqrt{x^2+y}} RR_1^* - \frac{\alpha_1 r_1 K_{2,1}}{[RR_1]_0\sqrt{x^2+y}} \tag{C.2.2.3}$$

We can show $\frac{dRR_1^*}{d[RR_1]_0} < 0$ by first assuming the opposite:

$$\frac{\alpha_1' K_{1,1} + \alpha_1 r_1 K_{2,1}}{[RR_1]_0\sqrt{x^2+y}} RR_1^* - \frac{\alpha_1 r_1 K_{2,1}}{[RR_1]_0\sqrt{x^2+y}} \geq 0$$
$$(\alpha_1' K_{1,1} + \alpha_1 r_1 K_{2,1}) RR_1^* - \alpha_1 r_1 K_{2,1} \geq 0$$
$$RR_1^* \geq \frac{\alpha_1 r_1 K_{2,1}}{\alpha_1' K_{1,1} + \alpha_1 r_1 K_{2,1}} \tag{C.2.2.4}$$

Now we need to see whether C.2.2.4 is true. To do this we notice we can separate out the $r_1\beta_1 - \beta_1'$ terms from C.1.1.16:

$$(r_1\beta_1 - \beta_1')((RR_1^*)^2 - RR_1^*) + (\alpha_1' K_{1,1} + \alpha_1 r_1 K_{2,1}) RR_1^* - \alpha_1 r_1 K_{2,1} = 0$$
$$(\alpha_1' K_{1,1} + \alpha_1 r_1 K_{2,1}) RR_1^* - \alpha_1 r_1 K_{2,1} = (r_1\beta_1 - \beta_1')(RR_1^* - (RR_1^*)^2) \tag{C.2.2.5}$$

We have previously shown that $r_1\beta_1 - \beta_1' < 0$ (C.1.2) and $RR_1^* \neq 1$ (C.1.3). From C.1.1.16 we can also conclude that $RR_1^* \neq 0$, as that would imply $\alpha_1 r_1 K_{2,1} = 0$, which is not possible as all three terms are positive. As such we can conclude that the right hand side of C.2.2.5 is negative, and can say:

$$(\alpha_1' K_{1,1} + \alpha_1 r_1 K_{2,1}) RR_1^* - \alpha_1 r_1 K_{2,1} < 0$$
$$RR_1^* < \frac{\alpha_1 r_1 K_{2,1}}{\alpha_1' K_{1,1} + \alpha_1 r_1 K_{2,1}} \tag{C.2.2.6}$$

By C.2.2.6 we can see that C.2.2.4 is impossible, allowing the conclusion that $\frac{dRR_1^*}{d[RR_1]_0} < 0$ when there are no competing substrates.

C.2.3 RR_1^* goes to 0 as $[RR_1]_0 \to \infty$

We can show that RR_1^* goes to 0 as $[RR_1]_0 \to \infty$ by first revisiting C.1.1.16 for a single RR:

$$(r_1\beta_1 - \beta_1')(RR_1^*)^2 - ((r_1\beta_1 - \beta_1') - (\alpha_1' K_{1,1} + \alpha_1 r_1 K_{2,1}))RR_1^* - \alpha_1 r_1 K_{2,1} = 0$$

Since $K_{1,1} \equiv K_{m,1,1}/[RR_1]_0$ and $K_{2,1} \equiv K_{m,2,1}/[RR_1]_0$, we can show:

$$\lim_{[RR_1]_0 \to \infty} K_{1,1} = \lim_{[RR_1]_0 \to \infty} K_{2,1} = 0 \tag{C.2.3.1}$$

We can use C.2.3.1 to simplify C.1.1.16 and solve for RR_1^* in the limit $[RR_1]_0 \to \infty$:

$$\begin{aligned} (r_1\beta_1 - \beta_1')\,(RR_1^*)^2 - (r_1\beta_1 - \beta_1')\,(RR_1^*) &= 0 \\ (RR_1^*)\,((RR_1^*) - 1) &= 0 \\ RR_1^* &= 0 \text{ or } 1 \end{aligned} \tag{C.2.3.2}$$

We know from subsection C.2.2 that RR_1^* is a strictly decreasing function of $[RR_1]_0$. We showed in subsection C.1.3 that $RR_1^* < 1$ for finite values of $[RR_1]_0$; since $[RR_1]_0$ must decrease, only $\lim_{[RR_1]_0 \to \infty} RR_1^* = 0$ is valid.

C.2.4 $\frac{d[RR_1^*]}{d[RR_1]_0}$ goes to 0 as $[RR_1]_0 \to \infty$

We can next show that $\frac{d[RR_1^*]}{d[RR_1]_0}$ goes to 0 as $[RR_1]_0 \to \infty$ by first revisiting C.2.1.3:

$$\frac{d[RR_1^*]}{d[RR_1]_0} = \frac{\frac{\alpha_1' r_1 K_{m,1,1}}{(\alpha_1' K_{m,1,1} + \beta_1'([RR_1]_0 - [RR_1^*]))^2}}{\frac{\alpha_1' r_1 K_{m,1,1}}{(\alpha_1' K_{m,1,1} + \beta_1'([RR_1]_0 - [RR_1^*]))^2} + \frac{\alpha_1 K_{m,2,1}}{(\alpha_1 K_{m,2,1} + \beta_1 [RR_1^*])^2}}$$

which we can simplify by defining:

$$a \equiv \alpha_1' r_1 K_{m,1,1} + \beta_1'([RR_1]_0 - [RR_1^*])$$

$$b \equiv \alpha_1 K_{m,2,1} + \beta_1 [RR_1^*]$$

$$c \equiv \alpha_1' r_1 K_{m,1,1}$$

$$d \equiv \alpha_1 K_{m,2,1}$$

Giving us:

$$\begin{aligned} \frac{d[RR_1^*]}{d[RR_1]_0} &= \frac{\frac{c}{a^2}}{\frac{c}{a^2} + \frac{d}{b^2}} \\ &= \frac{c \cdot b^2}{c \cdot b^2 + d \cdot a^2} \end{aligned} \tag{C.2.4.1}$$

If we multiply $c \cdot b^2$ by $1/([RR_1]_0)^2$ we get:

$$\begin{aligned} (c \cdot b^2) \cdot \frac{1}{([RR_1]_0)^2} &= c \cdot \left(\alpha_1 \frac{K_{m,2,1}}{[RR_1]_0} + \beta_1 \frac{[RR_1^*]}{[RR_1]_0} \right)^2 \\ &= c \cdot (\alpha_1 K_{2,1} + \beta_1 RR_1^*)^2 \end{aligned} \tag{C.2.4.2}$$

We can then take the limit of this expression as $[RR_1]_0 \to \infty$:

$$\lim_{[RR_1]_0 \to \infty} \frac{c \cdot b^2}{([RR_1]_0)^2} = c \cdot (0+0)^2 = 0 \tag{C.2.4.3}$$

If we multiply $d \cdot a^2$ by $1/([RR_1]_0)^2$ we get:

$$\begin{aligned} (d \cdot a^2) \cdot \frac{1}{([RR_1]_0)^2} &= d \cdot \left(\alpha_1' r_1 \frac{K_{m,1,1}}{[RR_1]_0} + \beta_1' - \beta_1' \frac{[RR_1^*]}{[RR_1]_0} \right)^2 \\ &= d \cdot (\alpha_1' r_1 K_{1,1} + \beta_1' - \beta_1' RR_1^*)^2 \end{aligned} \tag{C.2.4.4}$$

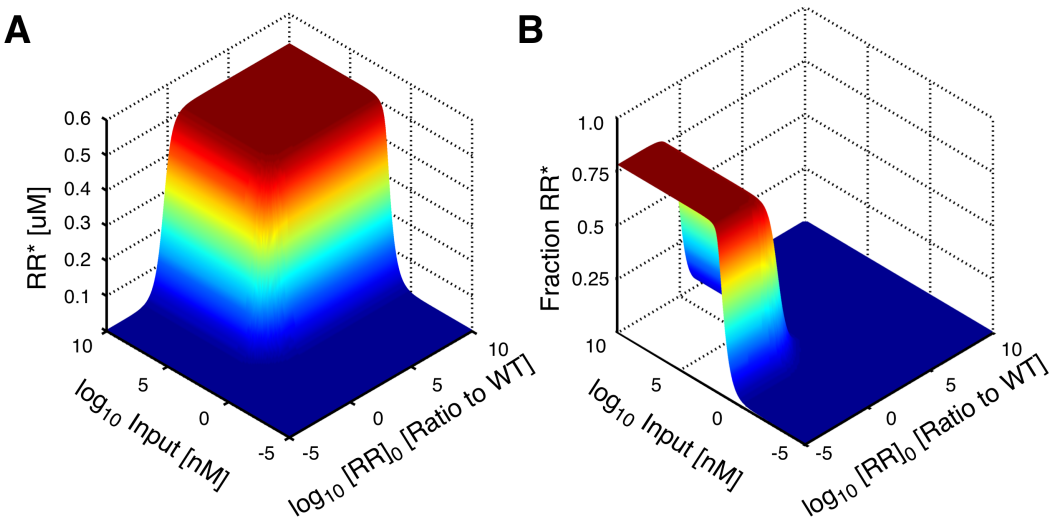

Figure C.2: Concentration vs. Fraction of Phosphorylated RR. The concentration (A) and fraction (B) of phosphorylated RR as a function of the concentration of Input and the total concentration $[\mathrm{RR}]_0$. The total concentration of RR is normalized by its wild-type concentration (1 μM). Both the Input concentration and $[\mathrm{RR}]_0$ are plotted on log scales.

We can then take the limit of this expression as $[RR_1]_0 \to \infty$:

$$\lim_{[RR_1]_0 \to \infty} \frac{d \cdot a^2}{([RR_1]_0)^2} = d \cdot (0 + \beta_1' + 0)^2 = d \cdot \beta_1'^2 \tag{C.2.4.5}$$

which is a constant. We then can use C.2.4.3 and C.2.4.5 to take the limit of C.2.4.1 as $[RR_1]_0 \to \infty$:

$$\lim_{[RR_1]_0 \to \infty} \left(\frac{c \cdot b^2}{c \cdot b^2 + d \cdot a^2} \left(\frac{[RR_1]_0}{[RR_1]_0} \right)^2 \right) = \frac{0}{0 + d \cdot \beta_1'^2} = 0 \tag{C.2.4.6}$$

Thus, although $[RR_1^*]$ generally increases as $[RR_1]_0$ increases, these increases eventually become very small. To demonstrate this change in behavior graphically, we used our more complex model including inputs and outputs (described fully in section C.3 below) to consider how $[RR_1^*]$ and RR_1^* change with increasing total RR concentration. As one can see from Fig. C.2, when $[RR_1]_0$ is high enough to saturate the HK, the maximum total concentration of active RR becomes insensitive to the amount of total RR in the system. Since the *concentration* of phosphorylated RR is constant,

the *fraction* of active *RR* molecules decreases as:

$$RR^* = \frac{[RR^*]}{[RR]_0} \sim \frac{1}{[RR]_0}$$

While TCS are thus certainly robust to changes in $[RR]_0$ at saturation [32], the system becomes progressively less efficient, in the sense that increases in total *RR* concentration do not produce concomitant increases in total possible *RR* activity. Note that the parameters used to produce Fig. C.2 are essentially all derived from experimental measurements (see section C.3). Interestingly, the measured *RR* concentration within cells (corresponding to $\log_{10}[RR]_0/[RR]_{\mathrm{WT}} = 0$ in Fig. C.2) is at a point that very nearly maximizes $[RR^*]$ while maintaining a relatively high fraction of phosphorylation.

C.3 Expanded Model With Inputs and Outputs

Bacteria can express many TCS pathways; as such we expanded the previously described model to include *N HK*s and *M RR*s, where each *HK* could interact with each *RR*. Since we also wanted to consider the effects of crosstalk on input and output functionalities, we included explicit inputs and outputs for each pathway. Consider a particular HK_i–RR_j pair from the set of all total *HK*'s

and *RR*'s. The set of reactions for this pair interacting with with input I_i is:

$$HK_i \underset{k_{b,i}}{\overset{k_{f,i}}{\rightleftharpoons}} HK_i^*$$

$$I_i + HK_i \underset{k_{-,i}}{\overset{k_{+,i}}{\rightleftharpoons}} I_iHK_i$$

$$I_i + HK_i^* \underset{k'_{-,i}}{\overset{k'_{+,i}}{\rightleftharpoons}} I_iHK_i^*$$

$$I_iHK_i \underset{k'_{b,i}}{\overset{k'_{f,i}}{\rightleftharpoons}} I_iHK_i^*$$

$$I_i + HK_iRR_j^* \underset{k_{-,i}}{\overset{k_{+,i}}{\rightleftharpoons}} I_iHK_iRR_j^*$$

$$I_i + HK_i^*RR_j \underset{k'_{-,i}}{\overset{k'_{+,i}}{\rightleftharpoons}} I_iHK_i^*RR_j$$

where the binding of input I_i is perfectly specific for HK_i. The binding of the input to the *HK* simulates the presence/absence of the environmental cue to which the *HK* responds. This input is able to bind to HK_i regardless of its phosphorylation and whether it has a bound *RR*. Note that the dissociation constant for *I* and *HK* does not depend on *RR* binding but could depend on the phosphorylation state of the *HK* in this model. A free *HK* molecule with a bound input experiences an increased rate of autophosphorylation, making the input an activator of the pathway (i.e. $k'_{f,i} > k_{f,i}$). The set of enzymatic reactions for the phosphorylation and dephosphorylation of RR_j by HK_i is:

$$HK_i^* + RR_j \underset{k_{-,1,i,j}}{\overset{k_{+,1,i,j}}{\rightleftharpoons}} HK_i^*RR_j \xrightarrow{k_{cat,1,i,j}} HK_i + RR_j^*$$

$$I_iHK_i^* + RR_j \underset{k_{-,1,i,j}}{\overset{k_{+,1,i,j}}{\rightleftharpoons}} I_iHK_i^*RR_j \xrightarrow{k_{cat,1,i,j}} I_iHK_i + RR_j^*$$

$$HK_i + RR_j^* \underset{k_{-,2,i,j}}{\overset{k_{+,2,i,j}}{\rightleftharpoons}} HK_iRR_j^* \xrightarrow{k_{cat,2,i,j}} HK_i + RR_j$$

$$I_iHK_i + RR_j^* \underset{k_{-,2,i,j}}{\overset{k_{+,2,i,j}}{\rightleftharpoons}} I_iHK_iRR_j^* \xrightarrow{k_{cat,2,i,j}} I_iHK_i + RR_j$$

In principle the model would allow for each of the $N \times M$ interactions to have independent rate parameters. None of the parameters for these reactions depend on whether the HK is bound to an I. Finally, we have the specific binding of output O_j by phosphorylated RR_j, producing the $RR_j^* O_j$:

$$RR_j^* + O_j \underset{k_{out,-,j}}{\overset{k_{out,+,j}}{\rightleftharpoons}} RR_j^* O_j$$

Since RRs are generally transcription factors, binding an output O simulates binding to and activating the transcription of DNA. As such, we consider the concentration of the $RR_j^* O_j$ complex to be the output activity of the pathway. The set of ODEs describing the concentrations of free HK_i and HK_i^* are:

$$\begin{aligned}
\frac{d[HK_i]}{dt} &= -\left(HK_i \cdot k_{f,i} + I_i \cdot HK_i \cdot k_{+,i} + \sum_j (HK_i \cdot RR_j^* \cdot k_{+,2,i,j}) \right) + \left(HK_i^* \cdot k_{b,i} + I_i HK_i \cdot k_{-,i} \right. \\
&\quad \left. + \sum_j (HK_i RR_j^* \cdot k_{-,2,i,j} + HK_i^* RR_j \cdot k_{cat,1,i,j} + HK_i RR_j^* \cdot k_{cat,2,i,j}) \right) \\
\frac{d[HK_i^*]}{dt} &= -\left(HK_i^* \cdot k_{b,i} + I_i \cdot HK_i^* \cdot k'_{+,i} + \sum_j (HK_i^* \cdot RR_j \cdot k_{+,1,i,j}) \right) + \left(HK_i \cdot k_{f,i} + I_i HK_i^* \cdot k'_{-,i} \right. \\
&\quad \left. + \sum_j (HK_i^* RR_j \cdot k_{-,1,i,j}) \right)
\end{aligned}$$

The set of ODEs describing HK_i bound by an input I_i are:

$$\begin{aligned}
\frac{d[I_i HK_i]}{dt} &= -\left(I_i HK_i \cdot k_{-,i} + I_i HK_i \cdot k'_{f,i} + \sum_j (I_i HK_i \cdot RR_j^* \cdot k_{+,2,i,j}) \right) + \left(I_i \cdot HK_i \cdot k_{+,i} \right. \\
&\quad \left. + I_i HK_i^* \cdot k'_{b,i} + \sum_j (I_i HK_i RR_j^* \cdot k_{-,2,i,j} + I_i HK_i^* RR_j \cdot k_{cat,1,i,j} + I_i HK_i RR_j^* \cdot k_{cat,2,i,j}) \right) \\
\frac{d[I_i HK_i^*]}{dt} &= -\left(I_i HK_i \cdot k'_{-,i} + I_i HK_i^* \cdot k'_{b,i} + \sum_j (I_i HK_i^* \cdot RR_j * k_{+,1,i,j}) \right) + \left(I_i \cdot HK_i^* \cdot k_{+,i} \right. \\
&\quad \left. + I_i HK_i \cdot k'_{f,i} + \sum_j (I_i HK_i^* RR_j \cdot k_{-,1,i,j}) \right)
\end{aligned}$$

Note that in the model there is a thermodynamic cycle describing the binding of each I to its cognate HK (see Fig. C.3). The parameters we used for the binding of input I_i by HK_i and the autophosphorylation/autodephosphorylation rates of HK_i are:

Parameter	Value
$k_{f,i}$	1 x 10^{-10} $\text{nM}^{-1}\text{s}^{-1}$
$k'_{f,i}$	0.1 $\text{nM}^{-1}\text{s}^{-1}$
$k_{b,i}$	0.1 s^{-1}
$k'_{b,i}$	0.1 s^{-1}
$k_{+,i}$	100 $\text{nM}^{-1}\text{s}^{-1}$
$k'_{+,i}$	100 $\text{nM}^{-1}\text{s}^{-1}$
$k_{-,i}$	1000 s^{-1}
$k'_{-,i}$	1 x 10^{-6} s^{-1}

For simplicity, we assumed HK's autodephosphorylation rate was not dependent upon the binding state of the input ($k'_{b,i} = k_{b,i}$). However, input binding greatly increases the autophosphorylation rate of the HK ($k'_{f,i} > k_{f,i}$). Additionally, the association rate of the input was assumed to not be dependent upon the phosphorylation state of the HK ($k'_{+,i} = k_{+,i}$), and the dissociation rate was chosen so that the system would not violate the Wegscheider condition (Fig. C.3):

$$k'_{-,i} = \frac{k_{-,i} \cdot k'_{b,i} \cdot k_{f,i} \cdot k'_{+,i}}{k_{+} \cdot k'_{f,i} \cdot k_{b,i}}$$

While detailed measurements of the parameters in the above table are not available, the interactions were chosen to represent biologically realistic values (i.e. the K_D of input binding is 10 nM, representing a sensor kinase that binds its input tightly).

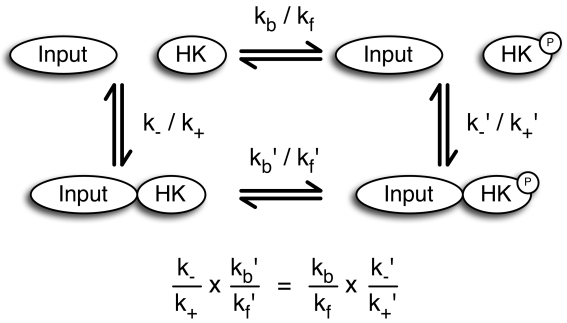

Figure C.3: Thermodynamic cycle for an input binding an HK. None of the parameters depend on whether the HK is bound to any of the M RRs. Note that the parameters are chosen so that the system does not violate the Wegscheider condition.

The set of ODEs describing the concentrations of free RR_j and RR_j^* are:

$$\begin{aligned}
\frac{d[RR_j]}{dt} &= \sum_i (-(HK_i^* \cdot RR_j \cdot k_{+,1,i,j} + I_i HK_i^* \cdot RR_j \cdot k_{+,1,i,j}) + (HK_i^* RR_j \cdot k_{-,1,i,j} + I_i HK_i^* RR_j \cdot k_{-,1,i,j} \\
&\quad + HK_i RR_j^* \cdot k_{cat,2,i,j} + I_i HK_i RR_j^* \cdot k_{cat,2,i,j})) \\
\frac{d[RR_j^*]}{dt} &= -\left(\sum_i (HK_i \cdot RR_j^* \cdot k_{+,2,i,j} + I_i HK_i \cdot RR_j^* \cdot k_{+,2,i,j}) + RR_j^* \cdot O_j \cdot k_{out,+,j} \right) \\
&\quad + \left(\sum_i (HK_i^* RR_j \cdot k_{cat,1,i,j} + I_i HK_i^* RR_j \cdot k_{cat,1,i,j} + HK_i RR_j^* \cdot k_{-,2,i,j} + I_i HK_i RR_j^* \cdot k_{-,2,i,j}) \right. \\
&\quad \left. + RR_j^* O_j \cdot k_{out,-,j} \right)
\end{aligned}$$

The set of ODEs describing the concentrations of bound *RR* are:

$$\begin{aligned}
\frac{d[HK_i^*RR_j]}{dt} &= -(HK_i^*RR_j \cdot k_{-,1,i,j} + HK_i^*RR_j \cdot k_{cat,1,i,j} + I_i \cdot HK_i^*RR_j \cdot k'_{+,i}) \\
&\quad + (HK_i^* \cdot RR_j \cdot k_{+,1,i,j} + I_i HK_i^*RR_j \cdot k'_{-,i}) \\
\frac{d[HK_iRR_j^*]}{dt} &= -(HK_iRR_j^* \cdot k_{-,2,i,j} + HK_iRR_j^* \cdot k_{cat,2,i,j} + I_i \cdot HK_iRR_j^* \cdot k_{+,i}) \\
&\quad + (HK_i \cdot RR_j^* \cdot k_{+,2,i,j} + I_i HK_iRR_j^* \cdot k_{-,i}) \\
\frac{d[I_iHK_i^*RR_j]}{dt} &= -(I_iHK_i^*RR_j \cdot k_{-,1,i,j} + I_iHK_i^*RR_j \cdot k_{cat,1,i,j} + I_iHK_i^*RR_j \cdot k'_{-,i}) \\
&\quad + (I_iHK_i^* \cdot RR_j \cdot k_{+,1,i,j} + I_i \cdot HK_i^*RR_j \cdot k'_{+,i}) \\
\frac{d[I_iHK_iRR_j^*]}{dt} &= -(I_iHK_iRR_j^* \cdot k_{-,2,i,j} + I_iHK_iRR_j^* \cdot k_{cat,2,i,j} + I_iHK_iRR_j^* \cdot k_{-,i}) \\
&\quad + (I_iHK_i \cdot RR_j^* \cdot k_{+,2,i,j} + I_i \cdot HK_iRR_j^* \cdot k_{+,i}) \\
\frac{d[RR_j^*O_j]}{dt} &= -(RR_j^*O_j \cdot k_{out,-,j}) + (RR_j^* \cdot O_j \cdot k_{out,+,j})
\end{aligned}$$

The set of ODEs describing the concentrations of free input *I* and output *O* are:

$$\begin{aligned}
\frac{d[I_i]}{dt} &= -\left(I_i \cdot HK_i \cdot k_{+,i} + I_i \cdot HK_i^* \cdot k'_{+,i} + \sum_j (I_i \cdot HK_iRR_j^* \cdot k_{+,i} + I_i \cdot HK_i^*RR_j \cdot k'_{+,i}) \right) \\
&\quad + \left(I_iHK_i \cdot k_{-,i} + I_iHK_i^* \cdot k'_{-,i} + \sum_j (I_iHK_iRR_j^* \cdot k_{-,i} + I_iHK_i^*RR_j \cdot k'_{-,i}) \right) \\
\frac{d[O_j]}{dt} &= -(RR_j^* \cdot O_j \cdot k_{out,+,j}) + (RR_j^*O_j \cdot k_{out,-,j})
\end{aligned}$$

A number of parameters relevant to *HK–RR* interactions have been directly measured. We were thus able to obtain the following values from the literature:

Parameter	Value	Reference
$K_{D,1,i,i} = k_{-,1,i,i}/k_{+,1,i,i}$	1 μM	[125]
$K_{D,2,i,i} = k_{-,2,i,i}/k_{+,2,i,i}$	1 μM	[125]
$k_{cat,1,i,j}$	1 s^{-1}	[5]
$k_{cat,2,i,j}$	0.1 s^{-1}	[3]
$[HK_i]_0$	100 nM	[32]
$[RR_j]_0$	1 μM	[32]

Note that the K_Ds listed above are for cognate interactions ($i = j$); the K_Ds for noncognate ($i \neq j$) interactions were generally varied in the model, depending on the specific simulation in question (see below). Based on these literature values, we set the following values for each of the rate constants in the model:

Parameter	Value
$k_{+,1,i,j}$	0.001 $nM^{-1}s^{-1}$
$k_{-,1,i,i}$	1 s^{-1}
$k_{-,1,i,j}$	Varies
$k_{cat,1,i,j}$	1 s^{-1}
$k_{+,2,i,j}$	0.001 $nM^{-1}s^{-1}$
$k_{-,2,i,i}$	1 s^{-1}
$k_{-,2,i,j}$	Varies
$k_{cat,2,i,j}$	0.1 s^{-1}
$k_{out,+,j}$	0.01 $nM^{-1}s^{-1}$
$k_{out,-,j}$	0.1 s^{-1}

As one can see from the above table, for simplicity we set all association rates, kinase catalytic rates, and phosphatase catalytic rates equal for all HK_i/RR_j pairs. Thus the only kinetic differences between cognate and noncognate interactions were the dissociation rates for the kinase and phosphatase interactions, which were set according to the particular noncognate K_D used in any given simulation (see below).

Finally, we used the following initial concentrations of the molecular species for most of our simulations:

Molecular Species	Initial Concentration
I_i	0 - 10^4 nM
HK_i	100 nM
RR_j	1 μM
O_j	100 nM

The total *HK* and *RR* concentrations were obtained from experimental measurements in bacterial cells (see the table above); the input concentrations were varied to obtain input-ouput curves (e.g. Fig. 4.1*D* in Chapter 4), and the output concentration was chosen to represent an intermediate concentration of *RR* binding sites ($\sim$10) within a bacterial genome. All of the remaining molecular species in the model had initial concentrations of 0.

C.3.1 Details on Parameters for Main Text Figures

For Fig. 4.1*C* of Chapter 4, we made a Goldbeter-Koshland model with a single input. The input interacted with the kinase *K* with the same schema and parameters as input binding of an *HK*, as described in the previous section. The phosphorylated *K* then could bind to an unphosphorylated substrate *S* and then transfer the phosphate group. An independent phosphatase *P* could then bind phosphorylated *S* and dephosphorylate it. The parameters for the kinase and phosphatase reactions with *S* are the same as described above for HK^* binding and phosphorylating *RR* and *HK* binding and dephosphorylating RR^*, respectively.

For Fig. 4.1*D* of Chapter 4, we used a model with a single *HK* and *RR* (i.e. $N = M = 1$) and set the total concentration of *RR* to be 100 nM (low concentration) and 10 μM (high concentration). In Fig. 4.2*A* we used a model with a single *HK* and two *RR*s ($N = 1$, $M = 2$). $[RR_1]_0$ was set to 1 μM and $[RR_2]_0$ was adjusted as indicated; in this case we used "cognate" K_Ds and kinetic parameters for both interactions. In Fig. 4.2*C* we used a model with an equal number of *HK*s and *RR*s (i.e. $N = M$); the total number of pathways was varied as indicated, and the noncognate

interactions ($HK_i - RR_j$ with $i \neq j$) had K_Ds assigned according to the indicated K_D ratio. We chose an input concentration in this case (12.35 nM) that resulted in half-maximal response for an isolated TCS pathway with cognate parameters. In Fig. 4.2*D* we once again used a model with a single *HK* and *RR*, starting with 2.5 μM phosphorylated HK^* and 2.5 μ M unphosphorylated *RR* and $[I]_0 = [O]_0 = 0$ in order to replicate the *in vitro* conditions considered by Skerker, *et al.* [28]. The cognate K_D in that case was set to the experimentally measured value (1 μM), and the noncognate interactions had K_Ds indicated by the specified K_D ratio. Finally, in the evolutionary trajectories discussed in Fig. 4.3, we considered a pathway with $N = M = 2$. For any given state along a trajectory, the presence of an arrow between two species (i.e. the existence of an interaction between a particular *I* and *HK*, or an *HK* and an *RR*) meant that the particular interaction had the "cognate" parameters indicated in the above tables. The absence of an arrow meant that the interaction had a K_D 10^6 times that indicated above; this fairly high K_D ratio was chosen to nearly complete absence of that interaction at that point along the trajectory.

Steady-state measurements were obtained by running the system until the levels of each species in the model stabilized. The actual time points at which the measurements were taken were chosen heuristically by visual inspection of the trajectories. The behaviors of systems containing a single *HK-RR* pair were confirmed analytically using our solution for RR^* (Fig. C.1).

C.4 Comparison of the Model with Experimental Data

In order to assess the validity of the model described in section C.3, we compared our predictions of pathway output at different HK and RR expressions with data from Batchelor and Goulian [32]. Replicating their experimental setup, we used the aforementioned "wild-type" parameters and protein concentrations and measured the fraction of bound output at various total concentrations of *RR* and *HK* (Fig. C.4*A* and *B*, respectively). We individually normalized the data from our simulations and from the experiments so that the Normalized Active Output at point *x* is calculated

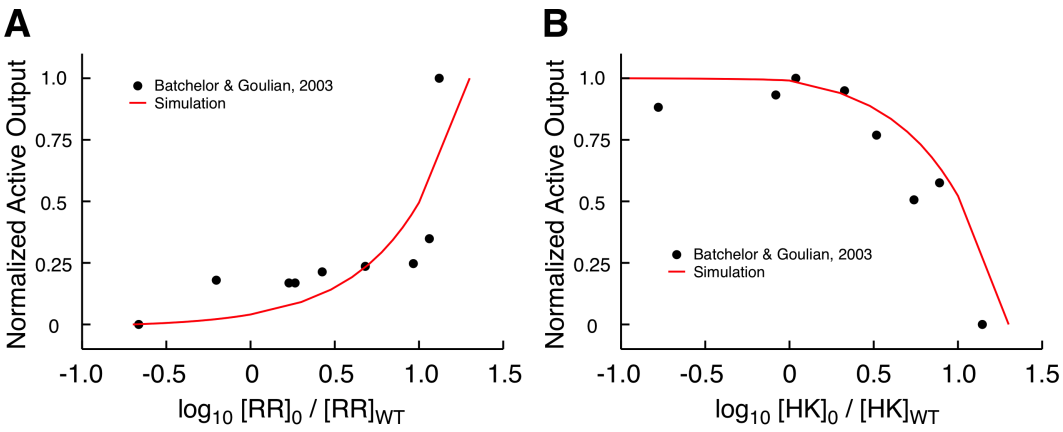

Figure C.4: Comparison of the model with experimental data from Batchelor & Goulian, 2003 [32]. The total concentrations $[\mathrm{RR}]_0$ (A) and $[\mathrm{HK}]_0$ (B) are normalized by their wild-type concentrations (1 μM and 100 nM, respectively) and are plotted on log scales.

as:

$$\text{Normalized Active Output}(x) = \frac{\text{Active Output}(x) - min(\text{Active Output})}{max(\text{Active Output}) - min(\text{Active Output})}$$

where x represents either a given $[RR]_0$ or $[HK]_0$ relative to the WT concentration. Note that we performed no fitting of the parameters for these simulations. We found that our model reproduces the general behaviors of the experimental data (Fig. C.4).

C.5 Details of K_A/K_S Analysis

C.5.1 Methods

The list of bacterial species was taken from the KEGG organism database [20]. In order to avoid redundancy across different strains of the same species, we selected only the first listed strain, resulting in 1267 bacterial genomes. Each of these genomes were searched for "histidine kinase" protein entries through the KEGG database keyword search; the UniProt ID, amino acid sequence and nucleotide sequence of each were saved. In order to assess the evolutionary dynamics of both the input functionality and crosstalk in HKs, we specifically focused on cases where the HK also

contained a "PAS" domain. PAS domains are the most common input domain in HK sequences, and also the most well-characterized experimentally. We obtained the PFAM domain annotations for each of these HK's and saved the locations of PAS sensor and histidine kinase domains for each HK sequence, where available [122]. We removed genomes that had less than 5 histidine kinases that each possessed an annotated PAS sensor and histidine kinase domain, resulting in 352 bacterial genomes. Additionally, we collected the UniProt ID, amino acid sequence and nucleotide sequence of every PAS domain-containing protein in each of the bacterial genomes. We also collected the UniProt ID, amino acid sequence and nucleotide sequence from all response regulators (RRs) in each of the 352 bacterial genomes.

We isolated the amino acid sequences for all domains in each HK based on the PFAM annotations. All PAS sensor and HK domain sequences from a given species were collected into separate files. Additionally, the HK domain sequence from Spo0B from *Bascillus subtilis* [9] and HK853 from *Thermotoga maritima* [27] were added independently to the HK domain sequence files. The domain sequence files were used as inputs for multiple sequence alignments using CLUSTALW [123]. The resulting amino acid alignments were used to map the codons from the associated nucleotide sequences. The alignments of the HK domains were also separated into two different sets: residues and their codons that align to contacting residues from either Spo0B or HK853 and those that do not [9]; we refer to these as the "K interface" and "K noninterface" residues, respectively. The RR sequences were aligned using CLUSTALW along with the sequence of Spo0F, the cognate response regulator of Spo0B [9]. We then similarly separated the RR alignments into two different sets: residues and their codons that align to contacting residues from Spo0F and those that do not [9]; we termed these as "RR interface" and "RR noninterface" residues.

These multiple sequence alignments at both the amino acid and codon level were used as inputs for calculating the K_A and K_S values for each pair using the *seqinr* library in R [124]. We specifically focused on the K_A and K_S values for the nearest neighbors in the alignment - domains that have most recently evolved from a single common ancestor. We identified these nearest neighbors through the phylogenetic trees obtained from the multiple sequence alignments; in this case the

nearest neighbors are existing domain sequences that have a single common ancestor in the tree for that genome's alignment. One issue with this data, however, is that Horizontal Gene Transfer (HGT) events might introduce "spurious" nearest neighbors, where the relationship between the two genes is *not* due to duplication and divergence of a common ancestor gene but rather due to the introduction of a new gene from a distantly related genome [121]. We performed an extensive analysis, detailed in the next section, in order to remove those cases from our data. The final datasets we used for our analysis included all the K_A and K_S values for these nearest neighbor domains across all species, so long as both K_A and K_S were not saturated and did not represent HGT events. Saturated pairs have experienced enough mutations that one cannot reliably estimate substitution rates, and are indicated by the *seqinr* library using K_S values of 10.

C.5.2 Horizontal Gene Transfer

New Two-Component pathways can enter a genome via two routes: duplication and divergence of an existing pathway (i.e. Lineage-Specific Expansion or LSE), or HGT from a separate lineage. As described above, we start with a pair of HK domain sequences that are nearest neighbors in the multiple sequence alignment of domains for a specific genome/species. The concern in this case is that HGT might introduce pairs of genes that, while nearest neighbors in the sequence alignment, are not related by duplication but rather due to HGT. Since our model only makes predictions regarding the evolutionary pressures that arise after duplication events, we must identify any "HGT pairs" in our data set and remove them from the analysis.

The prevalence of HGT among HKs was first estimated by Alm, *et al.* [121]. Following their approach, we first built a "species tree" for all of the species from which we obtained our HK sequences. To do this, we obtained the 15 genes Alm *et al.* used to generate their tree from each of our genomes, and built a multiple sequence alignment based on those genes. We then used the software package Phylip to generate a phylogeny from that alignment using 100 bootstrap replicates Felsenstein, *et al.* [161]. Using this species tree, we then built the phylogenetic profile for each gene, again following the approach of Alm *et al.* exactly [121]. The procedure itself

is summarized in Fig. C.5. We first generated a "reduced" species tree from the perspective of a given genome, in which taxa more distantly related to a given species are collapsed into a single "outgroup." A single HK sequence is chosen as a representative from each outgroup: this represents the HK gene from all the genomes that are in that outgroup that is "closest" to the HK gene in our starting genome based on Blastp score [162]. This set of representative sequences is then used to generate a profile for the gene, based on whether the genes from a given outgroup are closer in sequence to the starting gene than those from more distantly-related taxa (Fig. C.5B). In particular, if we look at some outgroup, we compare the Blastp score of the representative HK from that outgroup to the Blastp scores from all the other outgroups that are more distantly related to the original taxa.

Alm *et al.* define three different categories for each outgroup. An outgroup gets a "1" if the Blastp score from that outgroup is never more than 20 bitscore units worse than the Blastp scores from all the more distant outgroups. In that case, the idea is that the gene from that outgroup is closer than all the genes from outgroups that resulted from earlier branchings; or at least it is not much worse. This is what we expect if the gene is "present" in that genome in a way that is consistent with linear descent, and so it is assigned a 1. An outgroup is assigned a "0" if any more distant outgroup contains a gene that is more than 20 bitscore units closer to the original gene: this means that the gene is in a sense "absent," or at least that the pattern of sequence similarity is inconsistent with lineal descent. The only exception to this is if every genome in the outgroup in question contains less than 10 HKs; in that case, massive genome reduction may have resulted in the loss of the gene, and so the outgroup is assigned a "2" rather than a "0." In the work of Alm *et al.*, they generally considered any phylogenetic profile with more than two consecutive 0s to represent evidence of HGT.

We created phylogenetic profiles for each HK that was a member of a "nearest neighbor" pair with K_S less than 4 in our data set (i.e. the data that forms the basis of Figure 4 in the main text). We then looked for cases where either member of the pair appeared to be a "recent" HGT event; according to the definition of Alm *et al.*, these recent events focus on outgroups that are not

Table C.1: Table of coefficients for the linear fit of the log-log transformed K_A/K_S scatter plot from the interfacial residues of "K" domains' multiple sequence alignments

$$\log_{10} K_A/K_S = \beta_0 + \beta_1 \log_{10} K_S$$

Parameter	Value	p-value
$\hat{\beta}_0$	-0.46293	$< 2\text{x}10^{-16}$
$\hat{\beta}_1$	-0.86788	$< 2\text{x}10^{-16}$

Table C.2: Table of coefficients for the quadratic fit of the log-log transformed K_A/K_S scatter plot from the interfacial residues of "K" domains' multiple sequence alignments

$$\log_{10} K_A/K_S = \beta_0 + \beta_1 \log_{10} K_S + \beta_2 (\log_{10} K_S)^2$$

Parameter	Value	p-value
$\hat{\beta}_0$	-0.4577	$< 2\text{x}10^{-16}$
$\hat{\beta}_1$	-0.9923	$< 2\text{x}10^{-16}$
$\hat{\beta}_2$	-0.2683	0.163

too evolutionarily distant from the starting species in question. Of the 2,243 pairs we originally identified, 342 ($\sim$15%) represented HGT pairs according to this analysis, which we removed from our entire analysis.

C.5.3 Statistical Analyses

Once we obtained the K_A and K_S data for sequences that did not represent HGT events, we used R [88] to obtain regressions of the data to understand how the K_A/K_S ratio (i.e. the strength of positive or negative selection) varies over time for these domains. We used the synonymous substitution rate as a (rough) proxy for time since duplication, and as such we treated K_S as the independent variable for this analysis. First we performed a linear fit of the log-log transformed data from the interfacial residues of the "K" domains' multiple sequence alignments. We found that there was strong statistical significance for the coefficients (Table C.1). We then tested a quadratic fit of the data, but found that there was no statistical support for the quadratic term (Table C.2). We also performed a linear fit of the log-log transformed K_A/K_S data from the noninterfacial residues of the "K" domains' multiple sequence alignments. We found that the coefficients were statistically sig-

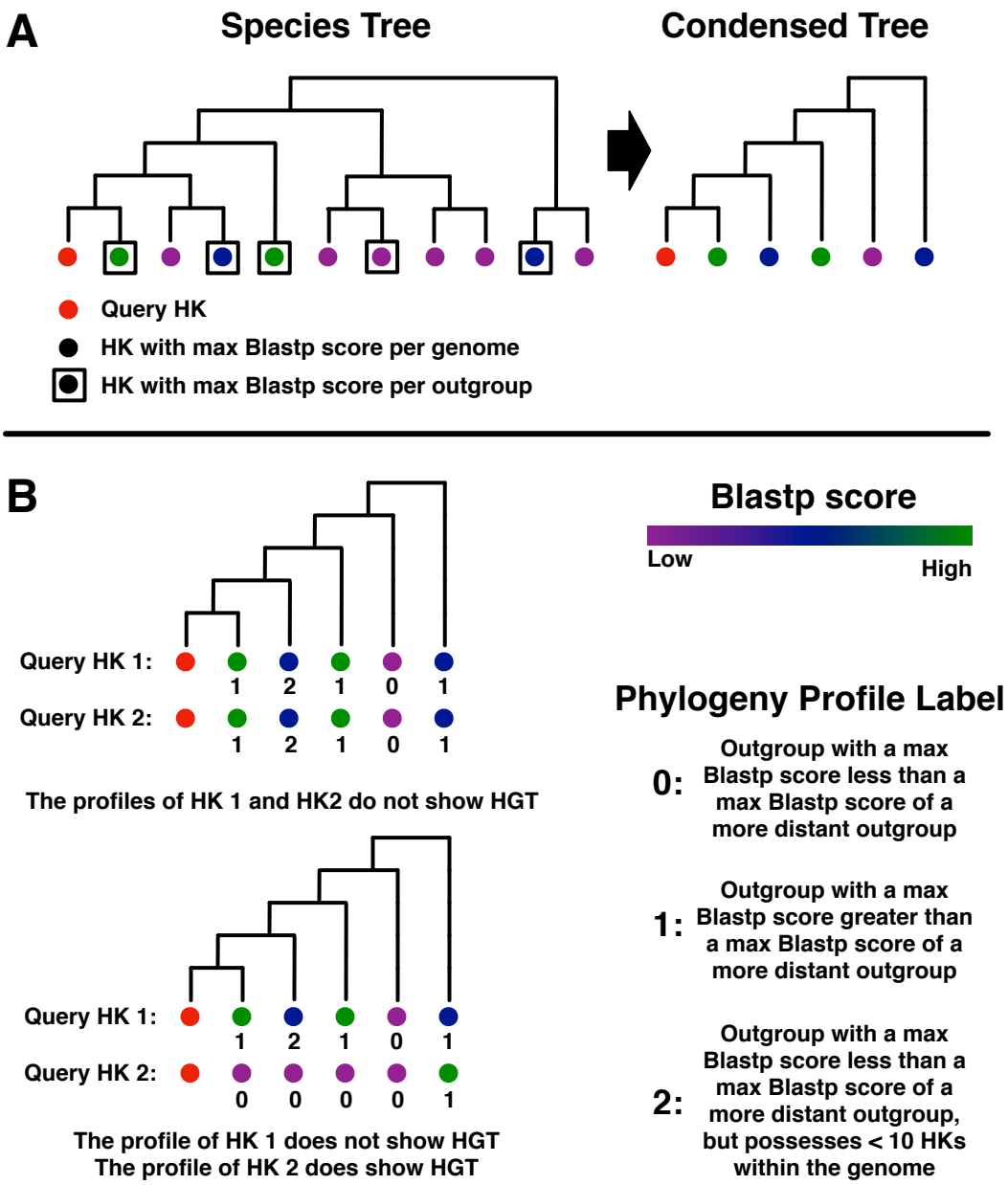

Figure C.5: (A) Generation of the condensed species tree. The highest Blastp bitscore from the genomes within an outgroup is taken as a representative of that taxa. For instance, the second outgroup from the Query HK (red) includes two genomes, one with a low Blastp bitscore, and another with a medium bitscore (purple and blue, respectively). This outgroup is collapsed to include just the medium bitscore HK. (B) Determination of the HK profiles. In the upper example HK 1 and HK 2 do not include consecutive 0s, indicating that neither come from an HGT. However, in the lower example, HK 2 includes a series of consecutive 0s, indicating that HK 2 was likely HGTed in from a distant branch of the tree.

Table C.3: Table of coefficients for the linear fit of the log-log transformed K_A/K_S scatter plot from the noninterfacial residues of the "K" domains' multiple sequence alignments.

$$\log_{10} K_A/K_S = \beta_0 + \beta_1 \log_{10} K_S$$

Parameter	Value	p-value
$\hat{\beta}_0$	-0.47611	$< 2\text{x}10^{-16}$
$\hat{\beta}_1$	-0.09997	$8.88\text{x}10^{-2}$

Table C.4: Table of coefficients for the quadratic fit of the log-log transformed K_A/K_S scatter plot from the noninterfacial residues of the "K" domains' multiple sequence alignments.

$$\log_{10} K_A/K_S = \beta_0 + \beta_1 \log_{10} K_S + \beta_2(\log_{10} K_S)^2$$

Parameter	Value	p-value
$\hat{\beta}_0$	-0.47391	$< 2\text{x}10^{-16}$
$\hat{\beta}_1$	-0.15215	0.113
$\hat{\beta}_2$	-0.11255	0.49

nificant, and indicated that the slope of this line was more shallow than that of the "K" interfacial residues (Table C.3). We also tried a quadratic fit of the data, but found no statistical significance for the quadratic term (Table C.4). We performed a further analysis to determine if the slopes of pure power-law regressions (i.e. linear fits of the log-transformed data) were significantly different between the interfacial and noninterfacial residue data sets. We combined the data sets and introduced a third categorical variable, Z. In this new dataset, $Z = 1$ when the entry in question is the "K" interface and $Z = 0$ when the entry is from the "K" noninterface. We then tested a combined linear model and found strong statistical support for the $\hat{\beta}_3$ estimate (Table C.5), indicating that the slopes in the two data sets are likely different from one another. Ultimately, this analysis suggests that the K interface undergoes rapid divergence immediately post duplication, while changes in the K noninterface occur more gradually over evolution.

The sizes of the aligned sequences of the interface and noninterface subsets of the K domains are significantly different, with the noninterface more than twice the length of the interface subset. In principle, K_A and K_S are substitution rates and as such they should be relatively insensitive to the size of the sequence used in the calculation; however, smaller sizes could certainly introduce

Table C.5: Table of coefficients for the linear model of the log-log transformed K_A/K_S scatter plot from the interfacial and noninterfacial residues of "K" domains' multiple sequence alignments. In this model, the parameter $Z = 1$ when K_S is from the "K" interface dataset and $Z = 0$ when K_S is from the "K" noninterface dataset.

$$\log_{10} K_A/K_S = \beta_0 + \beta_1 Z + \beta_2 \log_{10} K_S + \beta_3 Z \log_{10} K_S$$

Parameter	Value	p-value
$\hat{\beta_0}$	-0.47611	$< 2\text{x}10^{-16}$
$\hat{\beta_1}$	-0.09997	0.119
$\hat{\beta_2}$	0.01317	0.668
$\hat{\beta_3}$	-0.76791	$< 2\text{x}10^{-16}$

Table C.6: Table of coefficients for the linear fit of the log-log transformed K_A/K_S scatter plot from the the average of 10 random subsets from the noninterfacial residues of the "K" domains' multiple sequence alignments.

$$\log_{10} K_A/K_S = \beta_0 + \beta_1 \log_{10} K_S$$

Parameter	Value	p-value
$\hat{\beta_0}$	-0.27216	$< 2\text{x}10^{-16}$
$\hat{\beta_1}$	-0.02274	0.666

statistical biases in the estimation of K_A and K_S.

To control for the potential influence of this fact on our results, we calculated the the average substitution rates from 10 random subsets of the noninterfacial residues with the same total size as the interface subset. The results are shown in Fig. C.6. We performed linear and quadratic fits of the log-log transformed data from the noninterfacial datasets (Tables C.6 and C.7, respectively). We found no statistical support for either the linear or quadratic fits of the noninterfacial data. However, for display purposes, we plotted the linear fit in Fig. C.6. To determine if the slopes from these two datasets are statistically different, we combined the datasets with the third categorical variable $Z = 1$ when the entry is from the K interface and $Z = 0$ when the entry is from the random K noninterface subsets. We found strong statistical support for the $\hat{\beta_3}$ estimate (Table C.8), indicating the slopes in the two datasets are likely different from one another. This analysis suggests that the relative subset size does not strongly affect the overall trends in the data.

Additionally, protein folding constraints could strongly reduce amino acid substitution rates for

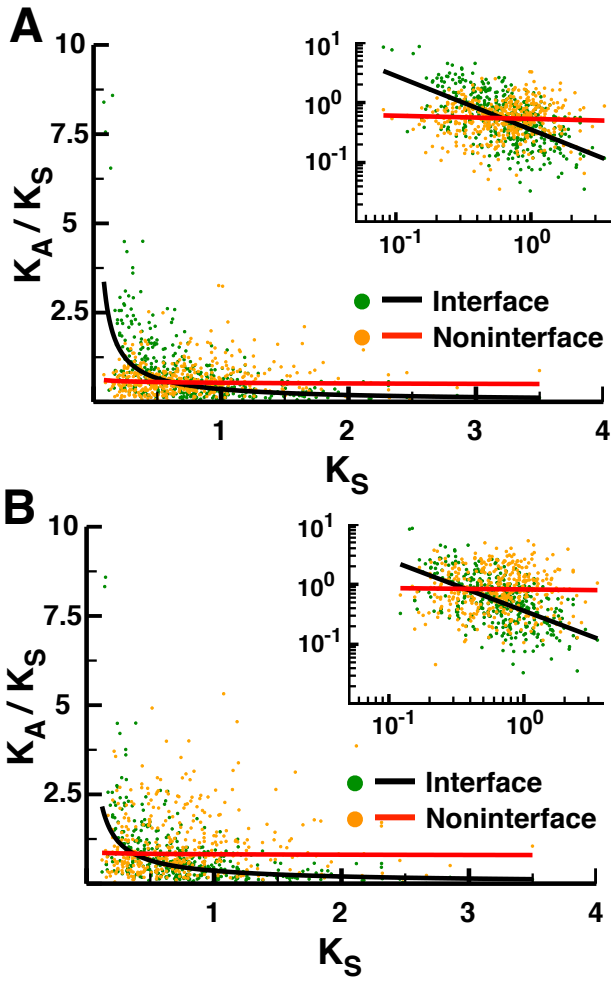

Figure C.6: (A) Comparison of the K_A/K_S vs. K_S for the K interfacial residues (green points and black curve) and the averages from 10 random subsets of the noninterfacial residues of the same total size (orange points and red curve). (B) Comparison of the K_A/K_S vs. K_S for the K interfacial residues (green points and black curve) and the averages from 10 random subsets of the noninterfacial **surface** residues of the same total size (orange points and red curve).

Table C.7: Table of coefficients for the quadratic fit of the log-log transformed K_A/K_S scatter plot from the average of 10 random subsets from the noninterfacial residues of the "K" domains' multiple sequence alignments.

$$\log_{10} K_A/K_S = \beta_0 + \beta_1 \log_{10} K_S + \beta_2 (\log_{10} K_S)^2$$

Parameter	Value	p-value
$\hat{\beta}_0$	-0.27052	$< 2\text{x}10^{-16}$
$\hat{\beta}_1$	-0.0147	0.862
$\hat{\beta}_2$	-0.08007	0.573

Table C.8: Table of coefficients for the linear model of the log-log transformed K_A/K_S scatter plot from the interfacial and the averages of 10 noninterfacial residues of "K" domains' multiple sequence alignments. In this model, the parameter $Z = 1$ when K_S is from the "K" interface dataset and $Z = 0$ when K_S is from the "K" noninterface dataset.

$$\log_{10} K_A/K_S = \beta_0 + \beta_1 Z + \beta_2 \log_{10} K_S + \beta_3 Z \log_{10} K_S$$

Parameter	Value	p-value
$\hat{\beta}_0$	-0.27216	$< 2\text{x}10^{-16}$
$\hat{\beta}_1$	-0.02274	0.688
$\hat{\beta}_2$	-0.17848	$9.03\text{x}10^{-11}$
$\hat{\beta}_3$	-0.91755	$< 2\text{x}10^{-16}$

Table C.9: Table of coefficients for the linear fit of the log-log transformed K_A/K_S scatter plot from the average of 10 random subsets from the noninterfacial surface residues of the "K" domains' multiple sequence alignments.

$$\log_{10} K_A/K_S = \beta_0 + \beta_1 \log_{10} K_S$$

Parameter	Value	p-value
$\hat{\beta}_0$	-0.0846	$3.0\text{x}10^{-4}$
$\hat{\beta}_1$	-0.02204	0.7519

buried residues. This could bias the K_A values for the noninterfacial residues of the K domains. As such we split the K domain residues into two subsets, those that are solvent exposed and those that are buried. To do this we calculated the Solvent Accessible Surface Area (SASA) of each residue in the Spo0B structure using POPS [163, 164]. Residues with 20% of their surface area exposed to water were defined as solvent exposed residues and the rest were considered buried.

We constructed 10 random subsets from the solvent-exposed noninterfacial residues of the same length as the interfacial residues subset and calculated the average K_A and K_S (Data shown in Fig. C.6B). There was, again, no statistical support for either the linear nor quadratic fits of the log-log transformed data (Tables C.9 and C.10). Additionally, we found that the slopes from the two datasets are likely to be different (Table C.11). As such, the overall trend of K_A being higher for the interface compared to the noninterface residues does not depend upon restricting the analysis to the surface residues. In the analysis described above, we used the multiple sequence alignments with Spo0B to define the interfacial residues of the K domains. We also performed

Table C.10: Table of coefficients for the quadratic fit of the log-log transformed K_A/K_S scatter plot from the average of 10 random subsets from the noninterfacial surface residues of the "K" domains' multiple sequence alignments.

$$\log_{10} K_A/K_S = \beta_0 + \beta_1 \log_{10} K_S + \beta_2(\log_{10} K_S)^2$$

Parameter	Value	p-value
$\hat{\beta}_0$	-0.07409	$2.0\text{x}10^{-3}$
$\hat{\beta}_1$	-0.13088	0.2275
$\hat{\beta}_2$	-0.36657	0.0663

Table C.11: Table of coefficients for the linear model of the log-log transformed K_A/K_S scatter plot from the interfacial and the averages of 10 noninterfacial surface residues of "K" domains' multiple sequence alignments. In this model, the parameter $Z = 1$ when K_S is from the "K" interface dataset and $Z = 0$ when K_S is from the "K" noninterface dataset.

$$\log_{10} K_A/K_S = \beta_0 + \beta_1 Z + \beta_2 \log_{10} K_S + \beta_3 Z \log_{10} K_S$$

Parameter	Value	p-value
$\hat{\beta}_0$	-0.0846	$1.63\text{x}10^{-4}$
$\hat{\beta}_1$	-0.02204	0.742628
$\hat{\beta}_2$	-0.35772	$< 2\text{x}10^{-16}$
$\hat{\beta}_3$	-0.8715	$< 2\text{x}10^{-16}$

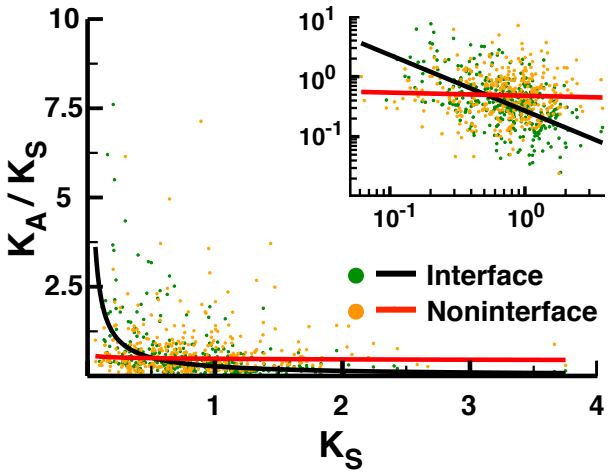

Figure C.7: Comparison of the K_A/K_S vs. K_S for the K interfacial residues (green points and black curve) and the noninterfacial residues (orange points and red curve), as identified by alignments to HK853 from *Thermotoga maritima*

similar analyses with HK853 from *Thermotoga maritima*, another histidine kinase with a solved crystal structure [27]. We aligned the K domain sequences and the primary sequence from the HK853 K domain and identified interfacial and noninterfacial amino acids and their corresponding codons. We then obtained K_A and K_S values for these subsets using *seqinr* library in R (Fig. C.7) [124]. We were able to obtain strong statistical support for the linear fit of the log-log transformed data for the interfacial residues, but did not find support for the quadratic term in the quadratic fit (Tables C.12 and C.13). We found no statistical support for either the linear nor quadratic fits of the log-log transformed data for the noninterfacial residues, although we use the linear fit for display purposes in Fig. C.7 (Tables C.14 and C.15). Finally, we found that the slopes from the two datasets are significantly different (Table C.16). As such, the overall trend of K_A for the interfacial residues being higher than that for the noninterfacial residues does not depend upon which protein is used to identify the interface. We then performed similar statistical analyses on the data from the interfacial and noninterfacial residues from the RR proteins. We found strong statistical support for the linear fit of the log-log transformed data (Table C.17). However, we did not get statistical support for the quadratic term in a quadratic fit of the transformed data (Table C.18).

We also performed a linear fit of the log-log transformed data from the noninterfacial residues of the RRs' multiple sequence alignments. We found no statistical support for the linear fit of

Table C.12: Table of coefficients for the linear fit of the log-log transformed K_A/K_S scatter plot from the interfacial surface residues of the "K" domains' multiple sequence alignments against HK853.

$$\log_{10} K_A/K_S = \beta_0 + \beta_1 \log_{10} K_S$$

Parameter	Value	p-value
$\hat{\beta}_0$	-0.56916	$< 2\text{x}10^{-16}$
$\hat{\beta}_1$	-0.93011	$< 2\text{x}10^{-16}$

Table C.13: Table of coefficients for the quadratic fit of the log-log transformed K_A/K_S scatter plot from the interfacial surface residues of the "K" domains' multiple sequence alignments against HK853.

$$\log_{10} K_A/K_S = \beta_0 + \beta_1 \log_{10} K_S + \beta_2 (\log_{10} K_S)^2$$

Parameter	Value	p-value
$\hat{\beta}_0$	-0.57286	$<2\text{x}10^{-16}$
$\hat{\beta}_1$	-0.86141	$5.67\text{x}10^{-15}$
$\hat{\beta}_2$	0.1342	0.397

Table C.14: Table of coefficients for the linear fit of the log-log transformed K_A/K_S scatter plot from the noninterfacial surface residues of the "K" domains' multiple sequence alignments against HK853.

$$\log_{10} K_A/K_S = \beta_0 + \beta_1 \log_{10} K_S$$

Parameter	Value	p-value
$\hat{\beta}_0$	-0.31771	$< 2\text{x}10^{-16}$
$\hat{\beta}_1$	-0.05164	0.418

Table C.15: Table of coefficients for the quadratic fit of the log-log transformed K_A/K_S scatter plot from the interfacial surface residues of the "K" domains' multiple sequence alignments against HK853.

$$\log_{10} K_A/K_S = \beta_0 + \beta_1 \log_{10} K_S + \beta_2 (\log_{10} K_S)^2$$

Parameter	Value	p-value
$\hat{\beta}_0$	-0.32036	$<2\text{x}10^{-16}$
$\hat{\beta}_1$	-0.002401	0.981
$\hat{\beta}_2$	0.096179	0.518

Table C.16: Table of coefficients for the linear model of the log-log transformed K_A/K_S scatter plot from the interfacial and the noninterfacial surface residues of "K" domains' multiple sequence alignments against HK853. In this model, the parameter $Z = 1$ when K_S is from the "K" interface dataset and $Z = 0$ when K_S is from the "K" noninterface dataset.

$$\log_{10} K_A/K_S = \beta_0 + \beta_1 Z + \beta_2 \log_{10} K_S + \beta_3 Z \log_{10} K_S$$

Parameter	Value	p-value
$\hat{\beta_0}$	-0.31771	$<$2x10^{-16}
$\hat{\beta_1}$	-0.05164	0.434
$\hat{\beta_2}$	-0.25145	1.63x10^{-15}
$\hat{\beta_3}$	-0.87847	$<$ 2x10^{-16}

Table C.17: Table of coefficients for the linear fit of the log-log transformed K_A/K_S scatter plot from the interfacial residues of the RRs' multiple sequence alignments.

$$\log_{10} K_A/K_S = \beta_0 + \beta_1 \log_{10} K_S$$

Parameter	Value	p-value
$\hat{\beta_0}$	-0.49457	$<$ 2x10^{-16}
$\hat{\beta_1}$	-0.7682	$<$ 2x10^{-16}

Table C.18: Table of coefficients for the quadratic fit of the log-log transformed K_A/K_S scatter plot from the interfacial residues of the RRs' multiple sequence alignments.

$$\log_{10} K_A/K_S = \beta_0 + \beta_1 \log_{10} K_S + \beta_2 (\log_{10} K_S)^2$$

Parameter	Value	p-value
$\hat{\beta_0}$	-0.48857	$<$ 2x10^{-16}
$\hat{\beta_1}$	-0.85732	5.55x10^{-12}
$\hat{\beta_2}$	-0.19359	0.334

Table C.19: Table of coefficients for the linear fit of the log-log transformed K_A/K_S scatter plot from the noninterfacial residues of the RRs' multiple sequence alignments.

$$\log_{10} K_A/K_S = \beta_0 + \beta_1 \log_{10} K_S$$

Parameter	Value	p-value
$\hat{\beta_0}$	-0.47316	$< 2\text{x}10^{-16}$
$\hat{\beta_1}$	-0.07323	0.167

Table C.20: Table of coefficients for the linear model of the log-log transformed K_A/K_S scatter plot from the interfacial and noninterfacial residues of RRs' multiple sequence alignments. In this model, the parameter $Z = 1$ when K_S is from the RR interface dataset and $Z = 0$ when K_S is from the RR noninterface dataset.

$$\log_{10} K_A/K_S = \beta_0 + \beta_1 Z + \beta_2 \log_{10} K_S + \beta_3 Z \log_{10} K_S$$

Parameter	Value	p-value
$\hat{\beta_0}$	-0.47316	$< 2\text{x}10^{-16}$
$\hat{\beta_1}$	-0.7323	0.268
$\hat{\beta_2}$	-0.02141	0.453
$\hat{\beta_3}$	-0.68497	$3.02\text{x}10^{-13}$

the log-log transformed data, indicating a lack of evidence for any strong correlation in the data (Table C.19). For visualization purposes, however, we included the linear fit in Fig. 4B of the main text for comparison to the interfacial residues. We also performed the analysis to determine if the slopes of the power-law regressions were significantly different between the RR interfacial and noninterfacial datasets as we had done with the HK. In the combined dataset the categorical variable $Z = 1$ when the entry is from the RR interface and $Z = 0$ when the entry is from the RR noninterface. We then tested a combined linear model and found strong statistical support for the $\hat{\beta_3}$ estimate (Table C.20), indicating that the slopes in the two data sets are likely different from one another. This once again suggests that the interfacial residues of the RRs undergoes rapid divergence post-duplication whereas the noninterfacial residues evolve slowly over a longer period of time. We then considered the possible effects of subset size and solvent-accessibility on the results from the RR datasets as we had done with the HKs. In this case we used the structure of Spo0F to determine the solvent accessible residues [165]. We performed the same analyses with

Table C.21: Table of coefficients for the linear fit of the log-log transformed K_A/K_S scatter plot from 10 random subsets from the noninterfacial residues of the RRs' multiple sequence alignments.

$$\log_{10} K_A/K_S = \beta_0 + \beta_1 \log_{10} K_S$$

Parameter	Value	p-value
$\hat{\beta_0}$	-0.26121	$< 2\text{x}10^{-16}$
$\hat{\beta_1}$	-0.01105	0.752

Table C.22: Table of coefficients for the linear model of the log-log transformed K_A/K_S scatter plot from the interfacial and 10 random subsets from the noninterfacial residues of RRs' multiple sequence alignments. In this model, the parameter $Z = 1$ when K_S is from the RR interface dataset and $Z = 0$ when K_S is from the RR noninterface dataset.

$$\log_{10} K_A/K_S = \beta_0 + \beta_1 Z + \beta_2 \log_{10} K_S + \beta_3 Z \log_{10} K_S$$

Parameter	Value	p-value
$\hat{\beta_0}$	-0.26121	$< 2\text{x}10^{-16}$
$\hat{\beta_1}$	-0.01105	0.812
$\hat{\beta_2}$	-0.16899	$1.39\text{x}10^{-15}$
$\hat{\beta_3}$	-0.77179	$< 2\text{x}10^{-16}$

the RR data as we had with the HKs (Tables C.21, C.23, C.22, C.24 and Fig. C.8). These analyses show that subset size and solvent-accessibility have no effect on the general trends shown in Fig. 4B of the main text. While the results described above strongly suggest an evolutionary pressure to diversify the K domain and RR interface residues after duplication events. To provide further evidence for this fact, we directly compared the K_A/K_S for the interface and non-interface residues of the K domain. As can be seen from Fig. C.9A, the majority of points in this case are *above* the line, indicating that the evolutionary pressure on the interface residues is higher than that on other

Table C.23: Table of coefficients for the linear fit of the log-log transformed K_A/K_S scatter plot from 10 random subsets from the noninterfacial surface residues of the RRs' multiple sequence alignments.

$$\log_{10} K_A/K_S = \beta_0 + \beta_1 \log_{10} K_S$$

Parameter	Value	p-value
$\hat{\beta_0}$	-0.25812	$< 2\text{x}10^{-16}$
$\hat{\beta_1}$	-0.08608	0.143

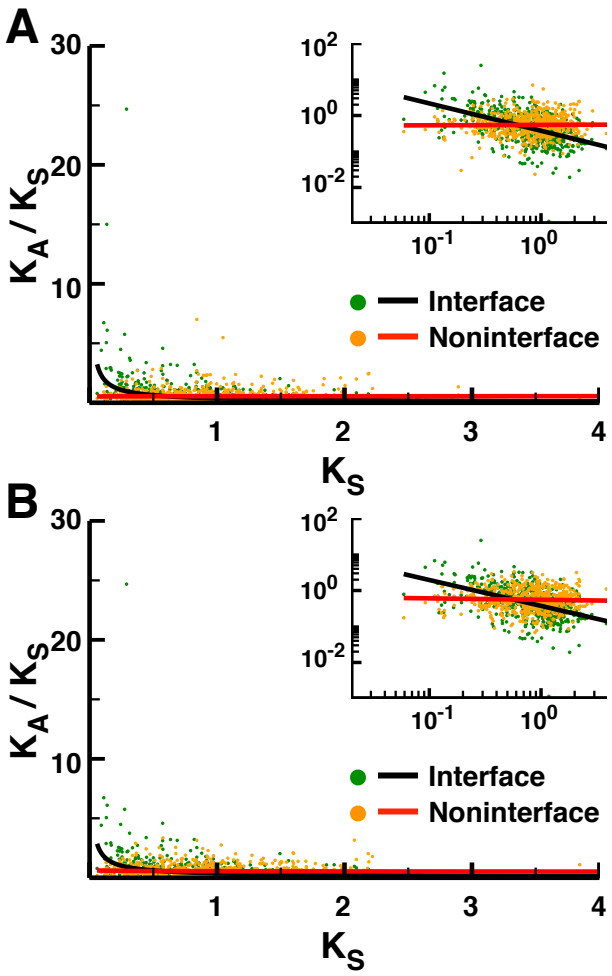

Figure C.8: (A) Comparison of the K_A/K_S vs. K_S for the RR interfacial residues (green points and black curve) and the averages from 10 random subsets of the noninterfacial residues of the same total size (orange points and red curve). (B) Comparison of the K_A/K_S vs. K_S for the RR interfacial residues (green points and black curve) and the averages from 10 random subsets of the noninterfacial **surface** residues of the same total size (orange points and red curve).

Table C.24: Table of coefficients for the linear model of the log-log transformed K_A/K_S scatter plot from the interfacial and 10 random subsets from the noninterfacial surface residues of RRs' multiple sequence alignments. In this model, the parameter $Z = 1$ when K_S is from the RR interface dataset and $Z = 0$ when K_S is from the RR noninterface dataset.

$$\log_{10} K_A/K_S = \beta_0 + \beta_1 Z + \beta_2 \log_{10} K_S + \beta_3 Z \log_{10} K_S$$

Parameter	Value	p-value
$\hat{\beta}_0$	-0.25812	$< 2\text{x}10^{-16}$
$\hat{\beta}_1$	-0.03958	0.424
$\hat{\beta}_2$	-0.16899	$3.25\text{x}10^{-15}$
$\hat{\beta}_3$	-0.67951	$< 2\text{x}10^{-16}$

residues in the sequence for most pairs in our data set ($p = 4.73 \times 10^{-9}$, binomial test). This is also true for the RR sequences (Fig. C.9B, $p = 5.44 \times 10^{-3}$, binomial test). We also considered raw substitution numbers in addition to estimates of evolutionary pressure. In this analysis, we simply counted the number of amino acid substitutions in a given subsequence; we call this substitution number n_A. For the plot in Fig. 4.4*D* of Chapter 4, we compared the number of substitutions in the K domain interface to the average number of substitutions in 10 random subsets of the same size. We found that the number of substitutions in the interface was, on average, much larger in the K domain interface than in the rest of the protein, even for recent duplications where the PAS domain seems to be older than the K domain (Figs. 4.4*C* and *D* in Chapter 4). To further analyze this data, we directly compared the substitution number in the K domain interface to the average obtained from the random subsets for each protein. As expected, we see a very strong statistical bias, with the bulk of points above the diagonal (Fig. C.10, $p < 2 \times 10^{-16}$, binomial test). This bias holds even if we consider only random subsets of residues that are on the surface but are not involved in the interface (Fig. C.10, $p < 2 \times 10^{-16}$, binomial test). The results from Fig. C.10 indicate that substitution rates in the interface are, on average, higher than they are for non-interface residues. To provide a concrete example of this phenomenon, we focused on a pair of very recently diverged proteins from the bacterium *Halococcus turkmenicus*. The alignment of the K domains from these two proteins (D2RXW7 and D2RTZ4) are shown, along with the alignment to the Spo0B sequence, in Fig. C.11. As one can see, in this case there are only a total of 7 substitutions (highlighted in red); of these, 4 are found in the interface region (indicated with asterisks). The substitution rate in the interface is thus $\sim 31\%$, compared to a substitution rate of $\sim 5\%$ among non-interface residues. The combined evidence from Fig. 4 in the main text, and the analyses presented above, indicate that the protein interaction interface between HKs and RRs is under significant diversifying evolutionary pressure after duplication, as our model predicts.

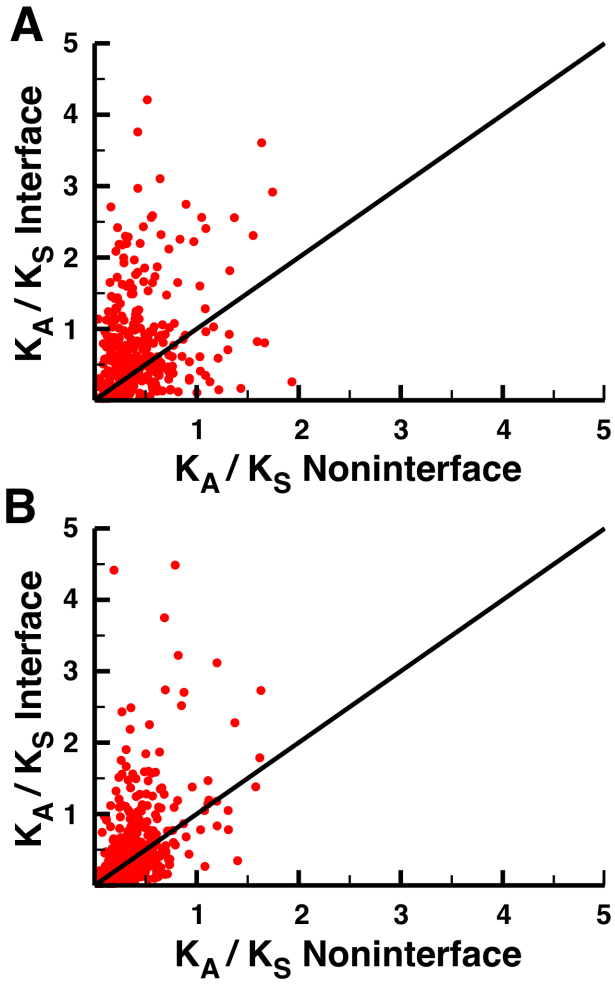

Figure C.9: (A) Comparison of the K_A/K_S values for the K domain interface and noninterface residues. Every point on the graph corresponds to a distinct non-HGT K domain pair. The K_A/K_S of the non-interface residues for that pair is shown on the x-axis, and the K_A/K_S values for the interface on the y-axis. (B) Similar to the plot in A, except in this case based on the RR interface vs. non-interface residues. Note that we have restricted the axes in both plots to $K_A/K_S < 5$ to show the bias in the data more clearly. All points in the data set (e.g. the points in Fig. 4 in the main text) were used to calculate p-values based on the binomial test.

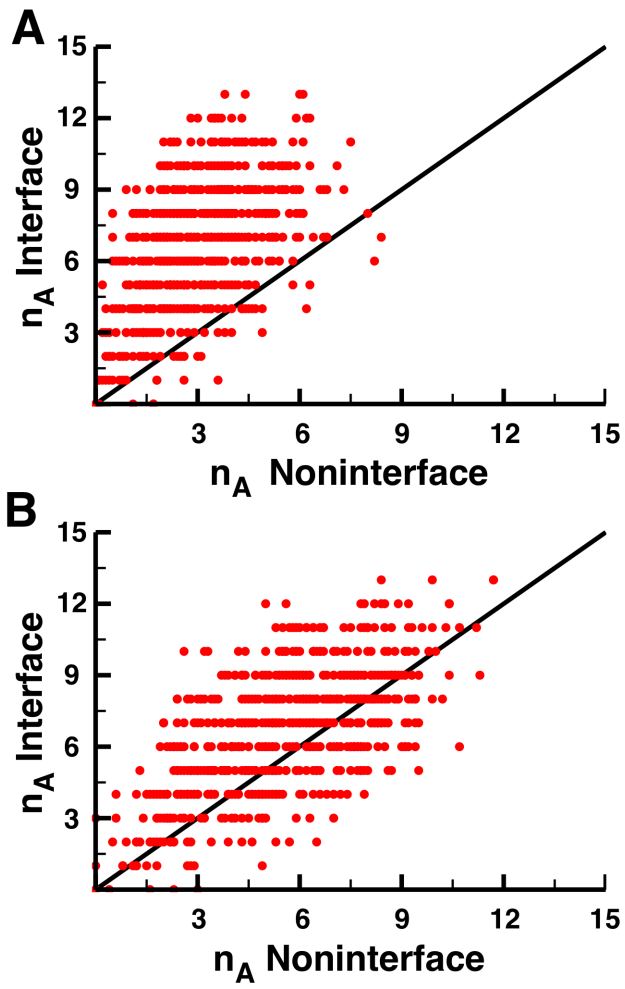

Figure C.10: (A) Comparison of the raw number of amino acid substitutions in the K domain interface vs. the average number in 10 random non-interface K domain subsets that have the same total number of residues. The amino acid substitution rate is clearly much higher in the interface in most cases. (B) As in A, but comparing the K domain interface to random subsets of **surface** residues. Again there is generally a considerably higher substitution rate in the interface residues.

```
D2RXW7      -RLEQFAYAASHDLQEPLRMVTSYLQLIENRYADE
D2RTZ4      -RLEQFAYAASHDLQEPLRMVSSYLQLIESRYTDV
Spo0b       ISDTALTNELIHLLGHSRHDWMNKLQLIKG----N
                           *      *  ** **    *

D2RXW7      LDADGREFIDFAVDGAERMREMIEGLLAYSRVETR
D2RTZ4      LDADGREFIDFAVDGAERMRDMIDGLLAYSRVETG
Spo0b       LSLQKYDRVFEMIEEMVIDAKHESKLSNLKTPH--
             ** *                *  **
```

Figure C.11: An example alignment for two close-related K domains in our data set, taken from the bacterium *Halococcus turkmenicus*. The alignment with the Spo0B sequence provides an indication of which residue positions are likely to be in the HK-RR interface; these positions are indicated with asterisks below the sequence. The amino acid substitutions between the two proteins are highlighted in red.

C.5.4 PAS domain analysis

In addition to predicting that there would be a strong evolutionary pressure to evolve new interface residues post-duplication, our model also predicted that any changes in input functionality should occur only after the interfaces had diverged (Figure 4.3 in Chapter 4). As mentioned above, we focused our analysis on the PAS domain, since it is a relatively common domain whose input functionality has been well established experimentally [121].

The analysis of PAS domain evolution in HKs is complicated by the fact that newly duplicated HK genes often do not evolve new PAS input domains *via* divergence of the PAS domain from the original sequence, but rather through the wholesale replacement of the PAS domain by swapping in the PAS domain from another protein. Indeed, in their analysis of HK evolution, Alm *et al.* found that this type of "domain swapping" was quite common [121].

To determine the extent of domain swapping in our data set, we took all the PAS domains from every genome we included in our analysis (including PAS domains from non-HK proteins) and performed a multiple sequence alignment of those domains from each genome. We then looked at each nearest-neighbor pair in our previous K domain alignment and asked whether the PAS domains from those sequences were also nearest neighbors in the PAS domain alignment. Interestingly, we could only find 67 ($\sim$3%) instances of this arrangement in our entire data set, indicating that domain swapping is by far more common than direct divergence for our HK pairs. Our analysis indicates considerably higher rates of domain swapping even compared to the work of Alm *et al.*; this discrepancy is likely due to the fact that we included *every* PAS domain from these genomes, rather than just PAS domains coming from other HK genes [121].

Our analysis thus indicated that there are only 67 clear instances where input functionality in a recently-diverged pair of HK sequences seems to have involved diversification of the original PAS domain. This represents too few sequences to preform a substitution-based analysis as described for the HK domains above. As such, we focused on understanding how the timing of swapping in a new PAS domain was related to the duplication of the HK gene. As described in the main text, there are two basic paths through which evolution of a novel HK gene with a swapped PAS domain

might occur. In one scenario ("scenario A"), an HK gene is duplicated and, after some period of time, a new PAS domain is "swapped in" from some other source. In the second scenario ("scenario B"), a gene that contains a PAS domain is duplicated, and then a new K domain is swapped in.

In our coarse-grained model of HK evolution, new inputs could evolve through divergence of the original PAS domain or through domain swapping; the model simply assumes that the input changes dramatically at some point, and is agnostic as to how this might occur. In the context of domain swapping, the model clearly predicts that scenario A should dominate, since changes in input functionality in this scenario would occur after the protein interaction interfaces have had sufficient time to diverge. To test this prediction, we calculated the K_S for each pair of K domains in our non-HGT set. For each of these pairs, we also calculated the K_S for the PAS from *each* HK gene, using the closest homologue of the PAS domain in question to calculate the K_S.

In Figure 4.4 of Chapter 4, we compare the K_S values for K domains against PAS domains. We find that the K_S of the K domains are significantly higher on average than the PAS domains (951/1300 above the $y = x$ line, $p < 2\text{x}10^{-16}$). In that analysis we did not differentiate PAS domains whose nearest neighbors are PAS domains from other HKs or from non-HK proteins. We separated this data into these two categories and reanalyzed the resulting datasets (Fig. C.12). We found that the K_S of the K domains was significantly higher on average than the PAS domains in both cases (711/956 above the $y = x$ line, $p < 2\text{x}10^{-16}$ and 240/344, $p = 1.632\text{x}10^{-13}$, respectively). As such, the overall trend found in Figure 4C of the main text does not depend upon the origin of the nearest neighbor of the PAS domain.

The biases seen in Fig. 4.4*C* and Fig. C.12 could represent the fact that it is simply more difficult to swap K domains than PAS domains. The fact that the bias holds when both genes are HKs, however (Fig. C.12*A*) indicates that this is likely not the case; there is no *a priori* reason that a new HK could not be formed first by duplication of one HK gene and subsequent swapping of a new K domain. The source of this bias is thus likely the fact that scenario B would involve creation of an HK that has different input functionality but still acts on a second RR due to crosstalk, which our model predicts would have serious fitness consequences for the cell.

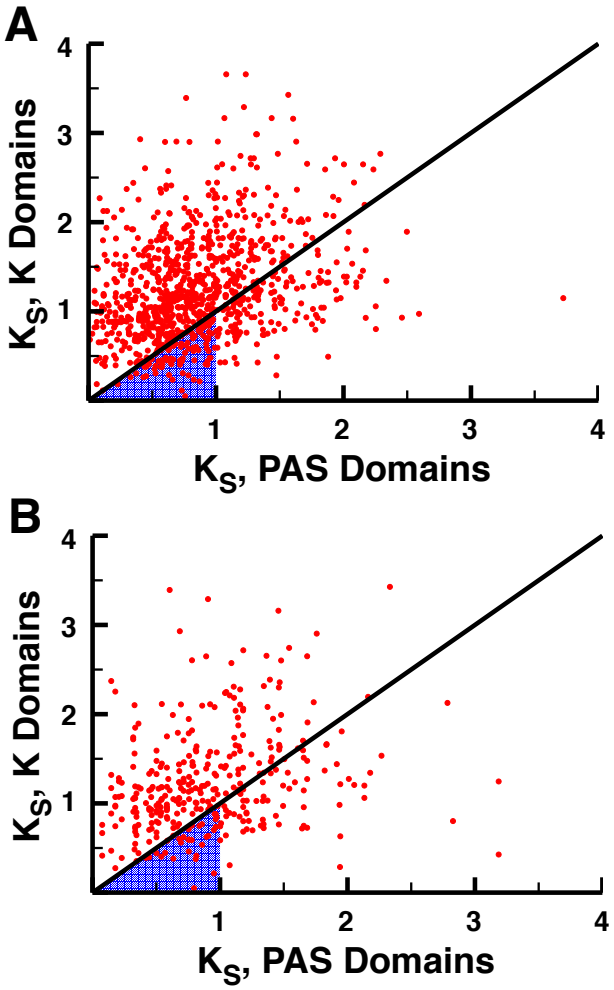

Figure C.12: (A) K_S of the K domains vs. K_S of the PAS domains in which the PAS domains are nearest neighbors with a PAS domain are PAS domains from other HKs. (B) K_S of the K domains vs. K_S of the PAS domains in which the PAS domains are nearest neighbors with a PAS domain are PAS domains from non-HK proteins.

Appendix D

Appendix for Crosstalk and the Evolvability of Intracellular Communications

D.1 Supporting Figures

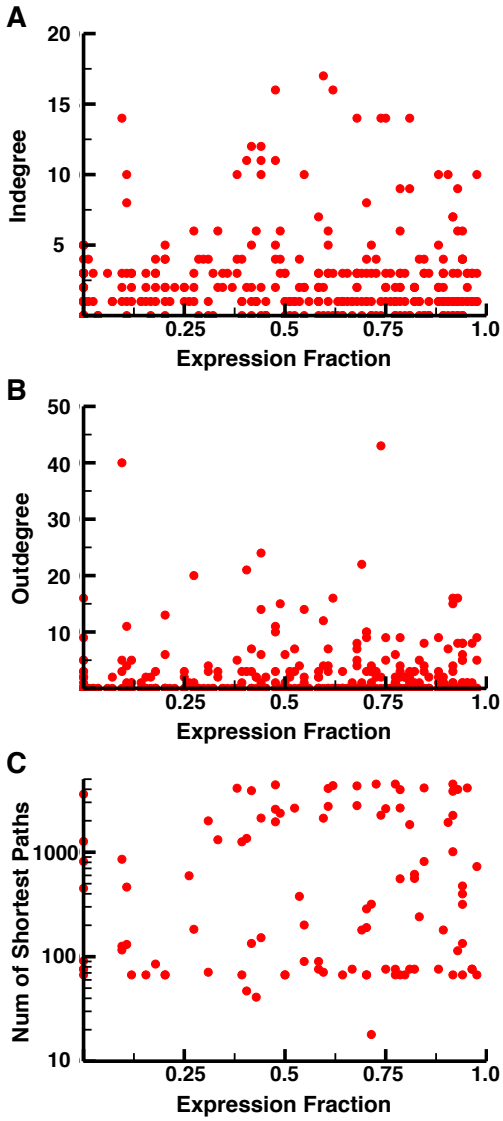

Figure D.1: Node properties versus the expression fraction of a node. (*A*) The indegree of a node versus its expression fraction. The expression fraction of a node is calculated as the ratio of the number of tissues in which the node is expressed to the total number of tissues. There is no correlation between the indegree and the expression fraction (Spearman's ρ = 0.03824749, p = 0.4795). (*B*) The outdegree of a node versus its expression fraction. There is a weak correlation between the outdegree and the expression fraction (Spearman's ρ = 0.15365, p = 4.285×10^{-3}), however the data cannot provide a statistically significant linear regression (p = 0.1548, adjusted R-squared = 0.003004). (*C*) The number of shortest paths between an input and an output that includes a node versus its expression fraction. There is a weak correlation between the number of shortest paths and the expression fraction (Spearman's ρ = 0.1506653, p = 5.106×10^{-3}), however the data cannot provide a statistically significant linear regression (p = 0.5059, adjusted R-squared = 0.005518).

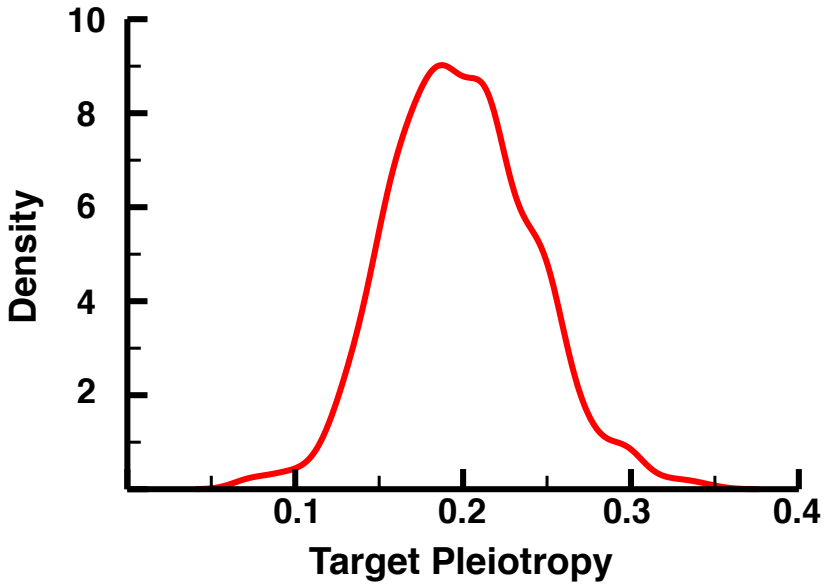

Figure D.2: Kernel density plot of target pleiotropy across the 84 expressed subnetworks. The target pleiotropy is the fraction of the 84 tissues in which inhibiting the node results in a change in output activity. Note that there are no nodes whose inhibition has an effect in all 84 tissues (Target pleiotropy = 1). Additionally, every node can be targeted in order to have an effect in at least one tissue. On average, an inhibitor will alter the output activity, and potentially change the phenotypic response of the cell, in about 17 of the 84 tissues (~20%).

D.2 Methods

D.2.1 Evolvable Boolean Networks

The evolvable Boolean networks start out with the same topology: two inputs (blue) that each activate their associated output (red) (Fig. D.3). These networks are randomly altered through one of three possible modifications: (1) adding an edge, (2) flipping an edge from activating to inhibiting or vice-versa, or (3) adding an intermediate node to connect two random nodes (See Fig. 5.2*A* of Chapter 5).

Following a modification, we run a synchronous Boolean simulation of the network for 100 steps with one, both or neither of the inputs active throughout the simulation. We then obtain the final activity of both outputs from each of the four simulations (Table D.1). These output activities are combined to form an 8 digit binary string, the 'I/O map' (i.e., the I/O map from Table D.1 would be '01110111').

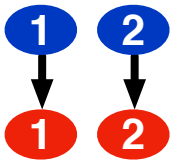

Figure D.3: Diagram of the initial TCS-like Boolean network

		Input 1 and Input 2 Activity			
		I_1I_2	I_1I_2	I_1I_2	I_1I_2
		00	10	01	11
Output Activity	O_1	0	1	0	1
	O_2	1	1	1	1

Table D.1: The final activity of outputs O_1 and O_2 for a modified network after a 100 step synchronous Boolean simulation. Each of the four rightmost columns represent the four different combinations of input activity, with neither input, input I_1, input I_2, or both inputs active throughout the simulation, respectively. From these results we would obtain the I/O map '01110111', which is the output O_1 and O_2 activities for each of the input combinations.

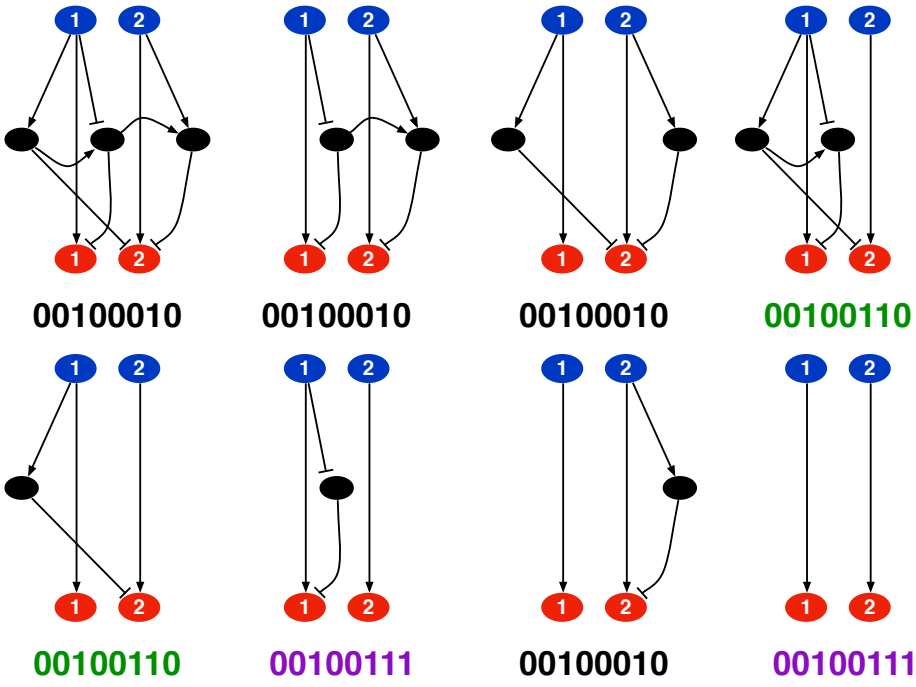

Figure D.4: Diagrams of each of 8 possible expression vectors for a hypothetical evolved network with three intermediate nodes. Beneath each subnetwork is the associated I/O Map, with different colors representing distinct maps. This example network thus presents 3 unique I/O Maps.

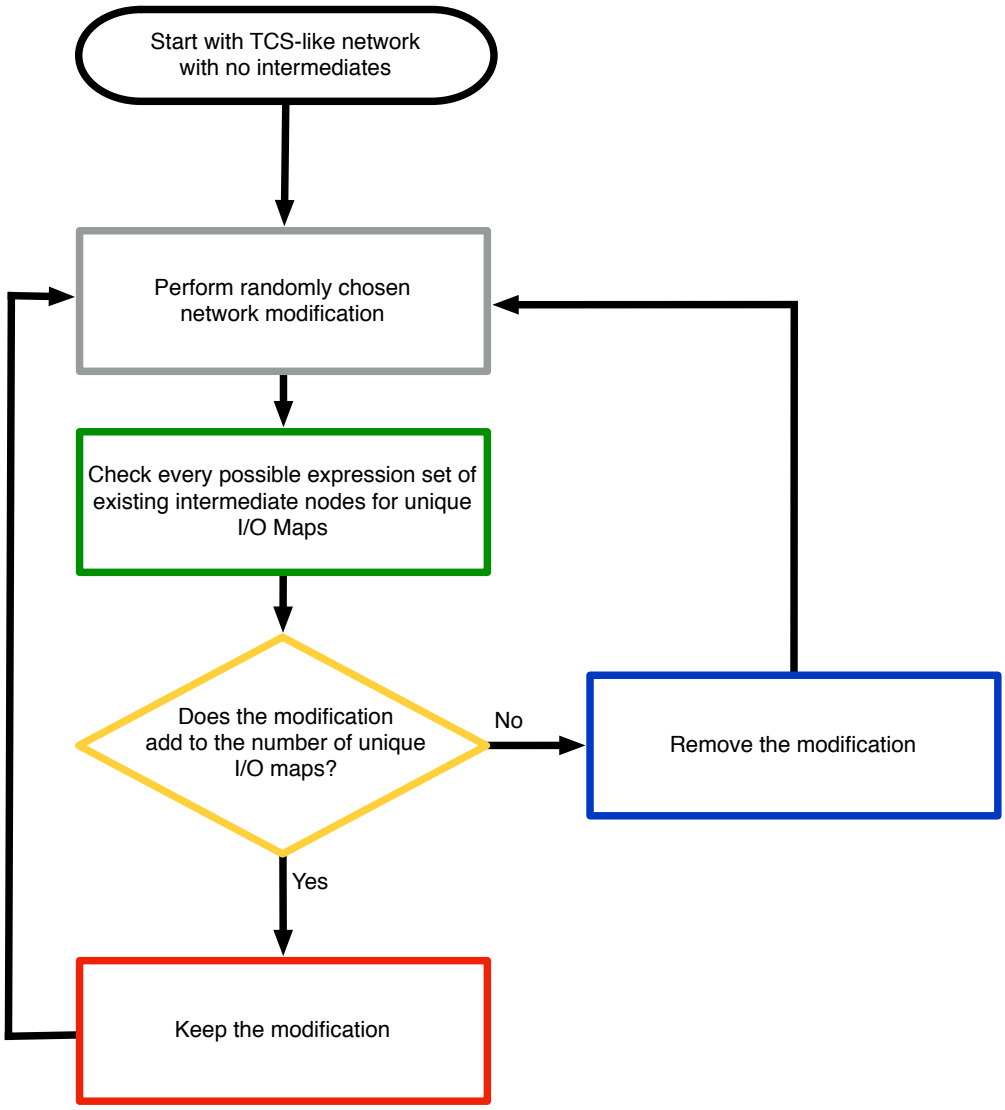

Figure D.5: The flow chart for the evolution of the Boolean signaling networks. After a modification is made to the existing network, the intermediate nodes are variably expressed in every possible combination. We then count the number of unique I/O maps and keep the modification if it increases the number of unique maps. Otherwise the modification is removed and the process is started again.

In order to explore the full potential of these systems, if a network has one or more intermediates, then we obtain the I/O map for every combination of expressed intermediate nodes, which we term an 'expression vector'. For example, a network with three intermediate nodes will have $2^3 = 8$ possible expression vectors, depending upon the presence or absence of each of the intermediates (see Fig. D.4). Due to the differences in network architecture, the network resulting from an expression vector has the potential to produce a unique I/O map. For a system with two inputs and two outputs, there is a maximum of 256 unique I/O maps possible. After each modification, we obtain the I/O map for each expression vector and count the number of unique I/O maps produced by the network. The number of unique maps is used to determine whether the modification is accepted: if the number of unique maps is increased by the modification, it is kept. New modifications are attempted until the network has 16 intermediate nodes. See Fig. D.5 for the flow chart of the evolutionary algorithm.

D.2.2 The Complete KEGG Signaling Network

We compiled the complete KEGG signaling network from the contents of the KGML files of 29 canonical pathways found in the KEGG Pathways database [20] (Table D.2). These pathways included node entries such as:

```
<entry id="6" name="hsa:4790 hsa:5970" type="gene"
        link="http://www.kegg.jp/dbget-bin/www_bget?hsa:4790+hsa:5970">
        <graphics name="NFKB1, EBP-1, KBF1, NF-kB1, NF-kappa-B,
            NF-kappaB,
              NFKB-p105, NFKB-p50, NFkappaB, p105, p50..."
              fgcolor="#000000" bgcolor="#BFFFBF"
              type="rectangle" x="984" y="311" width="46"
                  height="17"/>
    </entry>
```

where the 'name' attribute includes the KEGG entry for different isoforms of the same protein, in this case for NFκB1 and NFκB3. Each of these node entries are separated so that each KEGG entry becomes its own node, with each of these daughter nodes participating in the same incoming and outgoing edges. Once the nodes and edges from each of the canonical pathways were added, we collapsed the network by combining similar nodes. If two nodes had the same incoming and outgoing edges, these nodes were then combined into a single node. This process was done iteratively until there were no more nodes that could be combined.

For each of the nodes, we used the KEGG entries from each node to obtain the associated UniProt and ENSEMBL accession numbers. These were used to look up the expression data for the proteins included within each node from the Human Protein Atlas dataset [139]. The expression of any node in each of the tissues is dependent upon the expression of the genes associated the the node. If all of the associated genes are not expressed in a particular tissue, then the node is counted as not being expressed. However, if any of the genes are expressed, then the node is counted as being expressed in the tissue. This resulted in 84 expression vectors for the complete KEGG signaling network, generating 84 'expressed subnetworks', which represent the signaling network as it is expressed in each of the tissues. We then used the associated UniProt accession numbers to potentially annotate a node as an input (One UniProt entry includes the keyword 'Receptor') or as an output (One UniProt entry includes the keyword 'Transcription regulation') [89].

One of the metrics we use to compare subnetworks is the I/O map distance. To obtain the map distance we run each network in a synchronous Boolean simulation for 10000 steps with a set of N inputs active throughout the simulation. This is then done either for each of the inputs being activated individually (Fig. 5.3*C* of Chapter 5) or for 50 combinations of N inputs (Fig. 5.3*D* of Chapter 5). The activity of each output is averaged over the final 1000 steps to account for any oscillations in activity. This results in a matrix where each row is the average activity for all 67 outputs in response to a combination of active inputs. These respective elements in each of the matrices generated by the two subnetworks were then compared: the I/O map distance is the number of elements in the matrices that do not match.

Pathway	URL
Rap1	http://www.kegg.jp/kegg-bin/show_pathway?map=hsa04015
RAS	http://www.kegg.jp/kegg-bin/show_pathway?map=hsa04014
MAPK	http://www.kegg.jp/kegg-bin/show_pathway?map=hsa04010
ErbB	http://www.kegg.jp/kegg-bin/show_pathway?map=hsa04012
WNT	http://www.kegg.jp/kegg-bin/show_pathway?map=hsa04310
TGF-β	http://www.kegg.jp/kegg-bin/show_pathway?map=hsa04350
VEGF	http://www.kegg.jp/kegg-bin/show_pathway?map=hsa04370
TNF	http://www.kegg.jp/kegg-bin/show_pathway?map=hsa04668
HIF-1	http://www.kegg.jp/kegg-bin/show_pathway?map=hsa04066
NFκB	http://www.kegg.jp/kegg-bin/show_pathway?map=hsa04064
Jak-STAT	http://www.kegg.jp/kegg-bin/show_pathway?map=hsa04630
FoxO	http://www.kegg.jp/kegg-bin/show_pathway?map=hsa04068
Calcium	http://www.kegg.jp/kegg-bin/show_pathway?map=hsa04020
PI3K - Akt	http://www.kegg.jp/kegg-bin/show_pathway?map=hsa04151
mTOR	http://www.kegg.jp/kegg-bin/show_pathway?map=hsa04150
Toll-like Receptor	http://www.kegg.jp/kegg-bin/show_pathway?map=hsa04620
NOD-like Receptor	http://www.kegg.jp/kegg-bin/show_pathway?map=hsa04621
T-cell receptor	http://www.kegg.jp/kegg-bin/show_pathway?map=hsa04660
B-cell Receptor	http://www.kegg.jp/kegg-bin/show_pathway?map=hsa04662
Fcε RI	http://www.kegg.jp/kegg-bin/show_pathway?map=hsa04664
Fcγ R	http://www.kegg.jp/kegg-bin/show_pathway?map=hsa04666
Chemokine	http://www.kegg.jp/kegg-bin/show_pathway?map=hsa04062
Insulin	http://www.kegg.jp/kegg-bin/show_pathway?map=hsa04910
Adipocytokine	http://www.kegg.jp/kegg-bin/show_pathway?map=hsa04920
GnRH	http://www.kegg.jp/kegg-bin/show_pathway?map=hsa04912
Prolactin	http://www.kegg.jp/kegg-bin/show_pathway?map=hsa04917
Estrogen	http://www.kegg.jp/kegg-bin/show_pathway?map=hsa04915
Oxytocin	http://www.kegg.jp/kegg-bin/show_pathway?map=hsa04921
Neurotrophin	http://www.kegg.jp/kegg-bin/show_pathway?map=hsa04722

Table D.2: The list of 29 canonical pathways in the KEGG Pathways database that were compiled to create the complete signaling network

Appendix E

List of Publications

Rowland MA, Greenbaum J, Deeds EJ (2014) Crosstalk and the evolvability of intracellular communications. *Submitted*.

Rowland MA, Harrison B, Deeds EJ (2014) Phosphatase specificity and pathway insulation in signaling networks. *Submitted*.

Wani PS, **Rowland MA**, Odracek A, Deeds EJ, Roelofs J. (2014) Maturation of the proteasome core particle induces an affinity switch that controls regulatory particle association. *Submitted.*

Rowland MA and Deeds EJ. (2014) Crosstalk and the evolution of specificity in two-component signaling. Proc Natl Acad Sci USA. 111(15):5550-5.

Rowland MA, Fontana W, Deeds EJ. (2012) Crosstalk and competition in signaling networks. Biophys J. 103: 2389-2398.

*Featured in the Best of 2012 edition of *Biophysical Journal*

Appendix F

Permissions

Chapter 2 reproduced from:

RightsLink®

Account Info | Help | Live Chat

Title:	Crosstalk and Competition in Signaling Networks
Author:	Michael A. Rowland,Walter Fontana,Eric J. Deeds
Publication:	Biophysical Journal
Publisher:	Elsevier
Date:	Dec 5, 2012

Logged in as:
Michael Rowland
Account #:
3000816574

LOGOUT

Order Completed

Thank you very much for your order.

This is a License Agreement between Michael A Rowland ("You") and Elsevier ("Elsevier") The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

License number	Reference confirmation email for license number
License date	Jul 29, 2014
Licensed content publisher	Elsevier
Licensed content publication	Biophysical Journal
Licensed content title	Crosstalk and Competition in Signaling Networks
Licensed content author	Michael A. Rowland,Walter Fontana,Eric J. Deeds
Licensed content date	5 December 2012
Licensed content volume number	103
Licensed content issue number	11
Number of pages	10
Type of Use	reuse in a thesis/dissertation
Portion	full article
Format	both print and electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Title of your thesis/dissertation	Crosstalk, Network Dynamics, and the Evolution of Signaling
Expected completion date	Sep 2014
Elsevier VAT number	GB 494 6272 12
Billing Type	Invoice
Billing address	2030 Becker Drive LAWRENCE, KS 66045 United States
Permissions price	0.00 USD

VAT/Local Sales Tax	0.00 USD / 0.00 GBP
Total	0.00 USD

CLOSE WINDOW

8. Can I use my article in my dissertation or thesis without asking permission?

Yes, provided that you cite the original source and, for articles published in Volumes 90–105 (1993–2008), copyright notice.

Chapter 4 reproduced from:

Rowland MA and Deeds EJ. (2014) Crosstalk and the evolution of specificity in two-component signaling. Proc Natl Acad Sci USA. 111(15):5550-5.